

How to organize for the greenhouse

The sponsors of next week's conference at Geneva on climate change, the World Meteorological Organization and the UN Environmental Programme, should be careful not to push their luck by demanding too quick a treaty.

How does one seek to persuade the government of a country such as Uganda, where the incidence of AIDS among adults may exceed 10 per cent, that the prospect of global warming is an even more serious menace? Or a government such as that of the Soviet Union that economic growth and constitutional cohesion are less important than the avoidance of increased average surface temperature in other people's mid-latitudes? Or a government such as that of the United States, incapable on last week's evidence of charging its taxpayers for what they compel it by law to provide, that it should slap a tax on gasoline in the pursuit of atmospheric equilibrium when the same taxpayers reject the notion of an exactly similar tax targeted at economic survival? Evidently, a certain subtlety is required.

Subtlety, sadly, is not the strong suit of the organizers of next week's meeting at Geneva on climate change. The conference is almost calculated to evoke nothing tangible except howls of disappointment from the international environmentalist lobbyists who clamour for general adherence to a treaty for controlling greenhouse gases here and now. That is one reason why the organizers, the World Meteorological Organization (WMO) and the UN Environmental Programme (UNEP), should even at this late stage seek to exercise a calming influence on those attending. Another is that, because the excess greenhouse effect has indubitably come to stay, the more urgent need is to persuade constitutionally shortsighted governments to learn to look a century or so ahead. A third is that attempts to coerce governments into immediate action in causes of which they are not persuaded will usually leave them disaffected.

Yet at some stage there must be a treaty on greenhouse gases, preferably one to which all assent. All governments except those for the time being in crisis for other reasons agree with that.

So the trick that must be played on governments, in their own and their taxpayers' interests, is that they should start off with a common declaration of concern about global climate change — who in his senses could be unconcerned? — and that they should afterwards be compelled to make tangible (and monetary) commitments to the cause they have espoused. That is how the Montreal Protocol on ozone-unfriendly chemicals came about, with an initial declaration at Vienna. WMO and its friends are now saying that, with the precedent of Montreal, an inter-

national greenhouse convention could immediately spring into being. It would be more prudent to begin by collecting signatures for a general declaration. But next week is not the time for even that.

The meeting has a more specific purpose: to decide what to make of the reports of the three working groups of the Intergovernmental Panel on Climate Change (IPCC), which have been circulating in draft since June, to agree on the mechanism by which they should be rewritten and made to cohere with each other — and to keep IPCC in being. The first opportunity to sign a treaty will not come until February next year, at the intergovernmental conference arranged in Washington. The predictable protests that the Geneva meeting will have failed further to restrict the flow of motor cars will be beside the point.

There is nevertheless serious work to be done at Geneva. IPCC, like its three working groups, is a committee whose members are professionals, but which is not itself a professional organization. The result is that strong views flourish, but that the staff work is poorly done. Only that explains why IPCC's estimates of the consequences (called "impacts") of climate change, which anticipate by about a century the effects of carbon dioxide accumulation, have been put uncritically into circulation and uncritically accepted. It would help the greenhouse cause if IPCC could be reconstituted and only then renewed. The model should be UNSCEAR, the UN committee generating assessments of the hazards of ionizing radiation which is independent of its natural sponsor, the International Atomic Energy Agency. Getting WMO off IPCC's back would be a sufficient achievement for the Geneva meeting; finding more members carrying intellectual weight with the world at large, professional and otherwise, would be a bonus.

As things are, the outlook is not bright. Most of the influential governments at Geneva next week are likely to be preoccupied with other things — inflation, the Gulf, or both. They will be saying that it would have been better if the conference had been held a year ago, or if it could have been postponed. The only anchor is the diplomatic calendar; most of the people will turn up. Professionals' objectives at Geneva should be to reduce the greenhouse problem to its bare essentials, to get governments to agree that if there is a problem, something will have to be done. On such a platform, a convention could be built.



Too many precise and inevitably discordant estimates of how deeply the Maldives will be buried by the sea half a century from now will unfortunately have the opposite effect. There is already a remarkable degree of sympathy among governments for the concept of a treaty. The urgent need is to win the agreement of others (such as the United States) that a treaty of some kind is necessary. Only then will it be possible to decide what numbers should be entered. □

The Gallo case

The US National Institutes of Health should make public their apparently cheerful findings in their Gallo inquiry.

THE corrosive allegations against Dr Robert Gallo in respect of the discovery of HIV, the AIDS virus, have been substantially blunted by the inquiry carried out under the internal procedures of the National Institutes of Health (NIH), but they have not yet been entirely exorcized. That is the burden of the reports, last week (*Nature* 347, 502; 11 October 1990) and this (see page 603), that the NIH's quaintly named Office of Scientific Integrity has found no evidence to support earlier allegations that Gallo's HIV was from the outset, and knowingly, identical with that supplied to his laboratory by Dr Luc Montagnier at the Pasteur Institute and called "LAV".

The essence of NIH's finding is that there were other isolates of the virus growing in Gallo's laboratory, so that he had no need to misuse the samples sent from Paris. Gallo now follows the Office of Scientific Integrity in acknowledging that, for a virus distinguished by its lability, the strain of HIV grown and characterized at the National Cancer Institute at the end of 1983 and/or the beginning of 1984, and called "IIIB", is remarkably like LAV, and that accidental laboratory contamination may be the explanation. The continuing "formal investigation" at NIH is intended to discover whatever relationship there may be between LAV, IIIB and such relics as there are of the other virus strains known to have been in the laboratory at the time. The question naturally arises of whether this further investigation bears on the allegations levelled against Gallo, chiefly (but not exclusively) by the *Chicago Tribune* at the end of last year (see *Nature* 342, 462; 1989).

Naturally, it will be of more than casual interest if cells still stored in some refrigerator in Gallo's laboratory — known to have been contributed to the pool of infected cells from which virus producing cells were eventually isolated — can be made to yield virus with a nucleotide-sequence identical with that originally reported (*Nature* 313, 277; 1985). Then accidental contamination will be discounted — and the old puzzle of why two independent strains of such a labile virus should be so similar will be sharpened, as it has been already this year by the isolation from a patient in Nebraska of a virus very similar to LAV. (Even there, contamination cannot be excluded — see

Nature 347, 3; 1990.) Otherwise, there will be nothing to say, except that accidental contamination with LAV is a sufficient explanation.

Thus Gallo has much to gain and little to lose from the investigation now planned. In fairness to him, it should be completed. Yet there are equally good reasons why NIH should not wait on the outcome of a load of nucleotide-sequencing to make a public report on their inquiries so far. The allegations against Gallo were so serious, and so potentially damaging to the reputation to research, that only early publication of a measured report of the inquiry will lend persuasion to the information made public in the past two weeks. And while the Office of Scientific Integrity has certainly dealt with the central issue, peripheral issues not directly tackled may well be answered implicitly by a well-argued document. It is to be hoped that the supervisory committee of the US National Academy of Sciences will see the need for speed.

Can the whole business then be forgotten? By common consent, Gallo is a gifted and imaginative scientist who has contributed much to the deepening of understanding in the past two decades, not least by his discovery of interleukin-2. He is also a competitive, perhaps an over-competitive, scientist. It is marvellous that so much has been learned of the nature and the mechanism of AIDS in the decade since its recognition as a disease (in itself a remarkable achievement). The development in 1984 of a means of growing HIV continuously, leading to the diagnostic tests now in use, was a recognizably crucial step forward, even if the chronology of the discovery of HIV agreed between Gallo and Montagnier as part of the settlement of the Pasteur Institute's patent suit against the US government (*Nature* 326, 425; 1987) gave too little credit to the earlier work of Professors A. Karplus (Cambridge) and J. Levy (University of California, San Francisco), among others. The essence of his achievement, as of Montagnier's, is to have turned emergent understanding into a practical tool. Does it matter, in the circumstances, where the virus came from?

Much of the trouble of the past few years would not have arisen if Gallo had been able earlier to acknowledge that the similarity of IIIB and LAV required explanation, and that accidental contamination might possibly suffice. It is a pity that it has taken an inquiry by NIH to yield that now-evidently innocuous admission.

The diffidence may have sprung from fears that acknowledgement of the possibility of accidental contamination would prejudice the agreement between the French and US governments on the patents for the HIV diagnostic tests now in use, but that was probably always an unrealistic fear. Pride, a common human attribute (not always a failing) may have had a greater and more upsetting influence on relationships between two outstanding research groups cast by accident in competitive roles. It is to be hoped that, even before the further investigation is completed, that hatchet can be buried. Continuing argument about the rights and wrongs of the discovery of HIV will not, after all, do much to cure AIDS. □

End now in sight for Gallo investigation?

- NIH inquiry narrows its focus
- Doubts remain in Congress

Washington, Paris & London

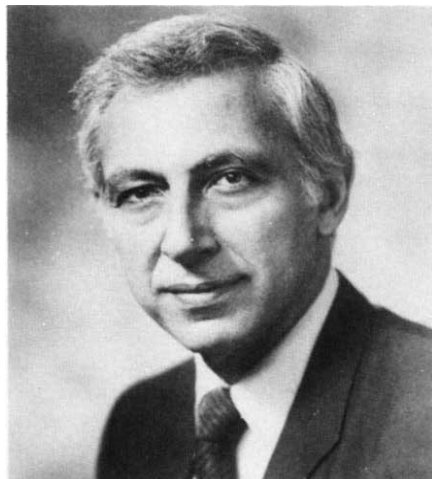
SOURCES close to the misconduct probe of prominent AIDS researcher Robert Gallo say that, with last week's announcement that officials at the National Institutes of Health (NIH) have found no substance to some of the most serious allegations of misconduct, the investigation is "virtually over".

Were NIH the only interested party, that might well be the case. But Congress, in the form of Representative John Dingell (Democrat, Michigan), the powerful government watchdog who has put pressure on NIH throughout the ten-month inquiry, may not be as easily satisfied. "Dingell will not stop this until all the questions are answered", says Peter Stockton, a staff member of Dingell's Energy and Commerce Investigations subcommittee. A meeting this week between NIH officials and subcommittee staff may decide whether Dingell will move from being an observer to a participant by holding hearings on the case.

One of the questions Dingell will want answered is why NIH have chosen to focus exclusively on issues surrounding two papers published in *Science* by Gallo's laboratory in May 1984 (see *Nature* 347, 502; 11 October 1990). In the *Chicago Tribune* article published last November that stimulated the NIH inquiry, reporter John Crewdson raised questions about Gallo's statements in the government's application for a patent on the first AIDS test, in correspondence with other researchers, and in public addresses, as well as in the *Science* papers.

William Raub, acting director of NIH, says investigators are "concentrating only on the reports in the refereed literature" to keep the probe "at the heart of what NIH does". "At this point we have decided not to concentrate on statements to the public", he says. If convincing evidence of misconduct arises from what is left of the investigation (mostly minor questions of wording and absent data in the papers), NIH may expand the probe, he says. "We don't expect that a major concentration of [other communications] will be useful at this point. We deliberately did not define the patent to be within the scope of the OSI investigation. Once the final report comes out, HHS [the Department of Health and Human Services, which runs NIH] may want to address that."

Raub stresses that although the charge of missing data is still worrying, "when one goes back 5-10 years one has to



Gallo's ordeal nearly over?

make allowance for a different attitude". Several recent misconduct cases have forced NIH to establish more stringent standards of record-keeping.

As for the most serious allegation that Gallo "stole" the virus from the Paris-based Institut Pasteur, Raub is satisfied for the moment. "The question of misappropriation is not on the table", he says, although he adds that it could reappear should new evidence materialize. "For those who have alleged he had nothing [at the time], the evidence is that he had other viruses. There is no obvious motive" for intentionally taking LAV, the French virus, he says, adding that the overall conclusions of the four seminal *Science* papers are not in doubt. In the remaining points of the investigation "we have an apparent mismatch between the work as described and the work as done", he says. "But none of that challenges that final outcome of the work."

Jean-Claude Chermann, the former director of the Institut Pasteur, says he is satisfied that Gallo "had enough isolates not to need LAV". Although he still has questions about Gallo's techniques, he is ready to see the dispute end. "I just want it to stop so we can get on with the science", he says. Luc Montagnier, discoverer of the French virus, describes the NIH investigation as an "American affair" and says he has nothing to add to the 1987 agreement he entered into with Gallo.

Without more data (Gallo's lawyer, Joseph Onek, says NIH have all the records from the laboratory during the key period), Raub says that it is "a fair presumption" that investigators may never know what happened in Gallo's laboratory in late 1983 and early 1984, the

key period before the *Science* papers were published.

At the moment, the most likely scenario for those months appears to be a story of contamination and confusion. Gallo concedes that, given that the French virus LAV and his own HTLV-IIIb are virtual genetic twins, it is possible that they were derived from the same sources. Although NIH investigators will try to trace the original viruses that were in his laboratory in 1983 and 1984, the odds that one of them should turn out to be an independently isolated copy of LAV appear slim (see *Nature* 347, 3 & 18; 6 September 1990).

That leaves laboratory contamination as the most likely source of the virus. Assuming that possibility, and given that Gallo says his chief virologist, Mikulas Popovik, kept almost no notes during the experiments, misstatements in published articles and public forums may have been inevitable. That those misstatements tended to favour a scenario in which HTLV-IIIb was an independent discovery of Gallo's laboratory may not be legitimate grounds for further investigation.

Bolstering the contamination theory is evidence that LAV has emerged unexpectedly in other laboratories around the world. Robin Weiss of the Institute of Cancer Research in London says that in 1985, when he was attempting to make new viral isolates in a laboratory where LAV was being grown, it was LAV that first emerged.

"In my opinion it's all over", says a source close to the investigation who has seen the laboratory data. "It could have been a contamination, but who cares?" Given the state of the science at the time, he says the distinction between the "continuous" growth of the virus in T cells that Gallo has claimed in the *Science* papers, and the intermittent growth that his records show, is academic. Gallo's laboratory "had 25 to 50 isolates growing", he says. "They weren't completely sustainable, but neither were anybody else's."

But even if NIH find no evidence of misconduct, and Dingell is satisfied that the investigation was conducted properly, broader questions raised in Crewdson's article and in the July issue of the New York-based satirical magazine *Spy* may still remain. Crewdson's and *Spy* author Seth Roberts's stories of backstabbing and deceit have painted a dim portrait of the AIDS research community. Roberts darkly hints that such activities have drawn attention from the very top. In a speech earlier this year, Pope John Paul II lashed out at "self-interested rivalries in the search for a medical answer" to the AIDS epidemic. But short of divine intervention, further investigation of AIDS researchers may be left to gossip magazines.

**Christopher Anderson, Peter Coles
& David Concar**

Cancer around nuclear plant

Boston

A US study released last week* shows an increase in the frequency of leukaemia among adults who lived and worked close to the Pilgrim nuclear power plant in Massachusetts in 1978–83. That news is certain to intensify the debate over links between cancer incidence and exposure to a low level of radiation, given that the study follows a US National Cancer Institute (NCI) study that found no increased incidence in cancer deaths around nuclear facilities and British data showing a cluster of cases of childhood leukaemia around the Sellafield nuclear plant (see *Nature* 347, 216 & 343, 679; 1990).

The Pilgrim study, carried out by the Massachusetts Department of Public Health, examined the incidence of leukaemia in persons living and working in 22 communities situated near the plant, taking into account how close to the plant they lived and worked and for how long. Although a fourfold incidence in leukaemia was found in people who lived within ten miles of the plant in the period 1979–83, no relationship between leukaemia cases and proximity to the plant was found during 1983–86. Leukaemia normally occurs at a rate of 2–3 cases per 100,000 population; the Pilgrim study was triggered by data suggesting that the rate close to the plant was double the Massachusetts average.

Assuming that it takes five years for leukaemia to develop, the 1979–83 increase in leukaemia incidence can be linked to a period of higher-than-normal emission of airborne radionuclides from the Pilgrim plant during the mid-1970s. Such emissions fell off considerably after 1978.

Three years ago, the Pilgrim plant was described by federal officials as “the worst-run” nuclear plant in the United States. But even during the periods of increased air emissions, the levels of radiation to which residents were exposed are believed to have been well within the Nuclear Regulatory Commission’s limit of 500 millirem per year. At least since 1980, exposure levels are believed to have been within the Environmental Protection Agency’s more stringent limit of 25 millirem per year.

Officials at Boston Edison, the utility that operates the Pilgrim plant, criticize the timing and methods of the study. But Daniel Hoffman, a researcher at the US Centers for Disease Control who led the committee that reviewed the results, called the study “well-designed, impressive, and surprising”. Hoffman stressed that

the Pilgrim study is one of the first to correct for confounding factors and says that it “suggests strongly that there may be increased risk levels [from radiation] at levels of exposure far lower than we would expect”.

Through interviews, the Pilgrim study collected detailed data about the 313 subjects involved and the length of time they had lived and worked near to the plant. The researchers tracked all types of adult leukaemia except chronic lymphocytic leukaemia (because it is widely considered to be the only type of leukaemia not associated with radiation). Confounding factors, including occupation, age, tobacco consumption and social status were taken into account, along with such factors as the proximity of toxic waste dumps and the origin of the drinking water. During the 1979–83 period, the researchers could even show that the likelihood of developing leukaemia was linked to where and

US SPACE PROGRAMME

Not so successful selling

Washington

If there remains any doubt that it is difficult to get private companies to enter the space market, a new report on commercialization at the National Aeronautics and Space Administration (NASA) is sure to remove it. The review*, by the Congressional General Accounting Office (GAO), tells the sad tale of seven projects (worth some \$700 million) that NASA tried to spin off to industry in 1989. After a year of solicitations with only one taker, six of the projects are now back at NASA, presumably delayed a year at substantial cost — casualties of the administration’s fervour for industry participation in the US space programme.

Ranging from a weightlessness laboratory to a robotic arm for the space station, the seven projects represented NASA’s best guess at candidates for commercialization in late 1988 when the space agency was in negotiations with the White House to reduce its 1990 budget request, the GAO says. As a compromise, NASA agreed to remove the seven projects from its budget, on the assumption that industry would take them over.

That assumption quickly proved false, GAO found. In five of the projects, “NASA found that commercial aerospace, construction, and finance companies were not willing to invest . . . because they perceived that few or no commercial markets existed”. (Several companies did, however, offer to loan

for how long people lived in the area.

Robert Knorr, one of the study’s two principal investigators, stressed the differences between the Pilgrim study and NCI’s recent mortality study. The Pilgrim study focused on a much smaller and more specific geographic area than did the NCI study, which drew only upon county data that would not detect highly-local increases in cancer. The Pilgrim study also looked at the incidence of disease rather than at only mortality.

After the results of the study were released, people living close to the plant flooded a state telephone ‘hotline’ number with calls. Massachusetts Governor Michael Dukakis has asked for the state to adopt the strictest radiation emission standard in the country, which would limit exposure of those living near nuclear plants to 10 millirem per year. State researchers have already begun studies on the incidence of childhood leukaemia in the vicinity of the Pilgrim plant and of leukaemia among workers at the plant.

Seth Shulman

NASA the money to build the projects itself.) GAO investigators also found that, in the case of three of the projects — a space station payload processing laboratory, an instrument processing laboratory and the robotic arm — the design process was as much as 60 per cent completed. To modify the facilities to allow for some commercial use would have required a costly and time-consuming redesign, prospective companies complained.

Among other errors, GAO found that NASA had badly miscalculated commercial financing costs, had underestimated the daunting technical risk of the projects, and had neglected to note that the safety requirements of certain projects limited the qualified private builders to those already working on the project under contract.

Except for an orbiting cryogenic pallet, NASA was forced to take all the projects back into its fold, although it was unable to find funding this year for four of them.

In part because of its experiences, NASA set up in July an internal commercial space steering committee. Chaired by NASA deputy administrator J. R. Thompson, the panel is expected to place high priority on finding agency projects that really are suitable for commercialization.

Christopher Anderson

Nature editorial changes

Rory Howlett has been appointed Book Review/Commentary Editor of *Nature*. Tim Lincoln is now News and Views Editor, and Maxine Clarke Executive Editor.

*Southeastern Massachusetts Health Study Final Report: Investigation of Leukemia Incidence in 22 Massachusetts Communities 1978–1986. Massachusetts Department of Public Health, Division of Environmental Health Assessment, October 1990.

*Space Projects: Improvements Needed in Selecting Future Projects for Private Financing. US General Accounting Office, September, 1990.

One man's achievement

Bodmin, Cornwall

FIVE winners of six Nobel prizes were among more than 100 scientists gathered in Bodmin last week to celebrate 25 years of a unique 'double experiment' at Peter Mitchell's Glynn Research Institute. At Glynn, Mitchell developed the theory of chemiosmosis (for which he won the Nobel prize for chemistry in 1978) and proved that a small independent research group with a talent for persuasive communication can have a disproportionate influence on its field of study.

Mitchell left the University of Edinburgh in 1963, plagued by stomach ulcers, to turn a derelict Cornish mansion into a research environment more sympathetic to what Mitchell describes as his "independent temperament" than the rigours of academic teaching and research. Joined by his former Edinburgh colleague Jennifer Moyle, Mitchell recruited a small team of like-minded researchers. From 1965, he refined and tested the chemiosmotic theory, which has since become a standard textbook theory in bioenergetics. Mitchell showed that oxidative phosphorylation of adenosine diphosphate in mitochondria to produce adenosine triphosphate, the 'energy currency' of the cell, depends on the movement of protons across mitochondrial membranes. Previously, biochemists had been searching for an energy-rich chemical intermediary to drive the process.

The replacement of this 'chemical coupling' theory with the idea of chemiosmosis as biochemical orthodoxy depended partly on Mitchell's concerted strategy of persuasion. From the late 1960s, sceptics were invited to Bodmin to discuss the conflicting theories. Mitchell says these meetings in Glynn's secluded

BIOTECHNOLOGY

Money no object

Washington

GENENTECH, Inc., in its first major project announcement since the completion of its \$2,100-million merger with the Swiss-based health-care conglomerate Roche Holdings Ltd, last week outlined plans to build a \$75-million research centre close to its main facility in South San Francisco. The merger agreement, which was completed on 7 September, provides Roche with a 60 per cent controlling interest in Genentech in return for the infusion of about \$490 million in capital by Roche (see *Nature* 343, 495; 1990). This sudden cash windfall has allowed Genentech to accelerate plans to expand its drug discovery research programme. The research centre, which is scheduled for completion by mid-1992, will provide research facilities for 420 people.

Diana Gershon

atmosphere were vital to resolve the "failure to communicate", that prevailed in the literature and during the hurly-burly of scientific conferences. The success of the strategy was documented on a campaign map (the 'Glynn Persuasion

European Commission and others) for its research funding. Mitchell has largely retired from research to become a full-time fund raiser, handing over to the present director of research, Peter Rich, recruited from the University of Cambridge in 1987.

Rich says that a single donation of several million pounds from some "eccentric

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Glynn House, Cornwall — a scientist's dream come true.

Monitor') with the leading proponents and opponents of the chemiosmotic theory around the world identified by colour-coded pins (green 'for'; red 'against'; and white 'on the fence').

Mitchell's success in his research, and in influencing his colleagues, is undoubted, but does the Glynn Institute provide a useful alternative model of how to run research at a time when science is becoming increasingly centred on huge collaborative efforts and dependent on central funding bodies? In a discussion of 'cost-effective research' at the jubilee meeting, Professor Harold Baum, from Kings College London, and a member of Glynn's governing council, argued that this is the case. Sir John Kendrew, winner of the 1962 Nobel prize for chemistry for his work on globular protein structure, lamented the current difficulty in getting grants to support original research. The assembled Nobel laureates, he said, had spent years "getting no results, other than bad ones" early in their research careers, and would not have been funded today.

Few would disagree that an independently funded group of fewer than ten researchers can provide a conducive atmosphere for original research. But even Glynn, headed by a Nobel prizewinner, must work hard to keep its head above water. The prospects for a series of Glynn-styled institutes would be bleak.

In its early days, Glynn's funding came from the sale of stock in a large construction company held by Mitchell and his brother. The 1978 Nobel prize cleared the bank overdraft, Mitchell says, but Glynn now relies on donations from wealthy individuals, charitable trusts and industry for its overhead costs, and central grant-making bodies (research councils, the

individual" and some astute investment would be sufficient to provide permanent funding for the institute. Until then, with reserves that would last for four years at the most, the struggle to maintain the supply of smaller donations is the priority. But Baum believes major benefactors will be forthcoming, most probably wealthy Japanese individuals, inspired by the "zen-like feeling" of the Glynn Institute.

Mitchell himself seems content simply to keep the Glynn Institute alive, rather than using its example to inspire the creation of similar research groups elsewhere. Glynn's methods are different from those in larger organizations, not necessarily better, he says. But he believes his own research success depended on Glynn's independence — if the institute was part of a larger organization, his research would have had to be "planned too far ahead", in order to justify continued funding.

Baum's enthusiasm for Glynn as a model for research organization is tempered somewhat by the realization that the institute is seen as an "aberration or artefact" by some sections of the scientific community. The concept of an independent laboratory run by a 'gentleman scientist' does seem more in tune with the nineteenth century than today's reliance on international collaboration in 'big science' projects; and few working scientists have the resources to follow Mitchell's example and set up their own institute. But as Roger Bourne, an Australian postdoctoral researcher recently arrived at Glynn, observes: "If a group of scientists sat down, got drunk, and fantasized about the type of laboratory they'd like to build — this would be it".

Peter Aldhous

Mouse patent a step closer

Munich

IN an important case for the European biotechnology industry, an appeals board at the European Patent Office (EPO) in Munich last week told patent examiners to reconsider their earlier decision that a transgenic mouse could not be patented. The mouse in question — the 'Oncomouse' — is produced by Du Pont Corporation and was patented successfully in the United States in 1988.

The Oncomouse carries human genes that increase its susceptibility to cancer and is widely used by cancer researchers. Other transgenic animals are already in use in medical and pharmaceutical laboratories and are expected eventually to revolutionize agriculture.

The original examiners had in June last year refused the patent on several grounds: first, that patents on "plant and animal varieties" are forbidden by Article 53(b) of the European Patent Convention (EPC); second, that the discovery had not been shown to be reproducible; and third, that patent law was not the appropriate tool for resolving the ethical questions raised by transgenic animals (see *Nature* 340, 85; 1989).

But the appeals board rejected the examiners' decision in all three instances. The board emphatically recommended that the examiners take up the issue of ethics and morality, which is specifically mentioned in Article 53(a) of EPC. According to Sandra Keegan, author of a European Commission directive concerning the patenting of animals and plants, the board is trying to amass a full range of arguments for and against the patent because the precedent is so important. The case will be a "tough nut to crack", says Keegan, especially since, according to an EPO spokesman, the examiners have a scientific and medical background and are by no means experts on ethics.

Du Pont argues that the market for transgenic animals in Europe will not develop without the same kind of patent protection that already covers marketing of the mouse in Europe. In the absence of a patent, it is trying to protect itself through licensing agreements.

In West Germany, environmentalists have objected vociferously to the idea of patenting animals because they consider it immoral and an affront to the natural order; animal breeders also object because they may be forced to pay licensing fees for each generation of animals that is derived from a patented animal. Last month, the Berlin-based *Gen-Ethisches Netzwerk* (gene-ethical network) advertised in the German press to try to raise money for a campaign against the patent, claiming that it would lead to exploitation of farmers in developing countries.

But Bob Beltz, a spokesman for Du Pont, says the activists have received more press coverage than their numbers warrant. "Don't underestimate the silent majority" who are in favour of animal patents, he says.

A European Commission proposal developed in the internal market directorate to increase patent protection for biotechnological discoveries including animals and plants has been bogged down for more than a year in the European Parliament. Keegan, who wrote the directive, says that although the issue is still an important one, the parliament has moved

AIDS TREATMENT

WHO concern over new drug

Munich & London

AN official of the World Health Organization (WHO) flew to Romania this week to investigate claims that an untried AIDS treatment is being administered to children in a Bucharest hospital. The WHO investigation of the Romanian trials comes on the heels of a separate inquiry into the effects of a 'wonder drug', known as Kemron, that emerged from Kenya (see *Nature* 347, 416; 1990).

Ever since AIDS became a problem, WHO has been overrun with letters and calls from "all types of people" who want to perform drug trials, says David Heymann, a specialist on AIDS at WHO.

"We don't want to deprive people of good AIDS drugs", says Heymann, "but we do want to see them be tested properly". The Romanian Health Ministry assured WHO by telephone last week that the drug used for the treatment, which is known as FLV 23/A, was harmless and that it would be tested in a double-blind trial using placebo controls under the supervision of physicians. But an official of WHO in Geneva said it would be impossible in the West to test a compound in humans before testing it in the laboratory. In the absence of published toxicity data, WHO was not able to say before the visit began this week if the drug had toxic side-effects.

The drug used in Romania was developed by David Hughes, a 61-year-old Scottish inventor and researcher. Hughes, in the company of several paediatricians, began earlier this month to administer the drug to about 100 HIV-positive babies, mostly orphans, and adults.

Romania is going ahead with the trials despite warnings from the British Foreign Office, which expressed doubts about Hughes' background and the claims he had made for the drug. Hughes claims to have a doctorate in "hyperbaric physiology" from the University of Grenoble,

on to other issues. Its unwillingness to act reflects a lack of consensus among European countries, which in her opinion will be resolved not through political discussion but rather through the revision of international conventions such as the EPC.

Although EPO has been friendly to patent applications involving biotechnology, it may take years before the mouse patent case is settled for good. The first European plant patent was awarded last year to the US company Lubrizol for a method to insert genes that boost the ability to store proteins in plants, and for the cells obtained using this method. A challenge to the patent is still pending.

Steven Dickman

according to the curriculum vitae Hughes presented to the Dubai-based International Medical Research Foundation (IMRF), which is paying for the research he is doing.

Nicolae Beldescu, director of the Romanian AIDS programme, defended the decision to go ahead with the treatments, saying that they had been approved by the National Drug Commission after toxicity tests on guinea pigs. Beldescu said that the commission concluded that it would be worth continuing the treatments "if only one baby is helped".

According to IMRF, the drug, which is based on oxides of hexylene based on cyclohexane, "will attack the AIDS virus and other similar viruses". Beldescu stressed that Hughes had never promised to "cure" AIDS patients, just to help them "recover". Western AIDS experts were not familiar with any successful treatment based on such a compound, which is not listed in drug reference books.

Hughes had tried the compound last year on AIDS patients in the main hospital in Lilongwe, the capital of the East African state of Malawi, but the positive results that he claims from the trials were negated by a lack of controls, say people familiar with the experiments.

A Western researcher who spent time in Malawi earlier this year, Lutz Gürtler of the University of Munich, said that before Hughes left Malawi physicians there had expressed suspicions about the legitimacy of his work.

Peter Desjardins, a board member of IMRF, says the trials, which also involve dietary restrictions, could not be properly run because some of the patients left the hospital at times. IMRF was founded in 1988 by private investors "largely" in order to support Hughes' research, says Desjardins.

Steven Dickman & Peter Aldhous

Stakes high in tax debate

Washington

IN the debris of the US federal budget, which Congress struggled to rebuild last week in an eleventh-hour compromise, is a little-known but hard-fought clause that could produce some \$1,500 million in extra research spending over the next five years.

Known as the small business research and experimentation (R&E) tax credit, the provision is an important element in the White House's plan for spurring industrial growth and reversing the country's economic decline.

Combined with the existing R&E tax credit for companies of all kinds, the proposed small business R&E credit (which increases the tax concession from 20 per cent to 30 per cent for companies worth less than \$50 million) could boost industrial research by more than \$20,000 million between 1991 and 1995, according to a recent study* by the Brookings Institution, a Washington-based independent think-tank. But because the credits would also cost the federal government some \$7,400 million in revenue over that period (no small matter in the face of a \$300,000 million deficit), their future remains uncertain as an austerity-minded Congress scrambles to make budget cuts.

The research tax credit has found both strong friends and strong enemies. Backing the credits is the Republican administration, which is betting that revenue sacrifices over the next five years will pay off in handsome dividends later, when the products of the additional research hit the market. But congressional critics point out that similar research credits have delivered precious little since their introduction in 1981.

The latest official analysis of the tax concessions, a study completed late last year by the General Accounting Office (GAO), found that the first five years of R&E credits stimulated less than \$2,500 million in new research at a cost of \$7,000 million in revenue lost to the government.

Since the GAO report, however, the picture has changed considerably. One of the chief problems with the original R&E credits was that they were based on the increase in research spending from one year to the next. Because Congress was reluctant to give any industry a permanent free ride, a company was allowed to recover only a percentage of that portion of the research that exceeded the average spending over the previous three-year period. Of course, come the next year, the increased spending had raised the three-year average, so businesses were forced to boost research once again to obtain the credit.

Eventually, many companies found that

even a 25 per cent reimbursement was not sufficient to justify spending more on research. "You could run yourself into the poor house trying to keep taking advantage of the credit", recalls Christopher Hill, executive director of the National Academy of Engineering's manufacturing forum.

In the 1989 budget reconciliation bill, however, Congress decided to abandon the 'moving average' in favour of a fixed base percentage, calculated as the ratio of research spending to sales between the years 1984 and 1988. Each year, companies could recover 20 per cent of the research that exceeded a level determined by multiplying the average of the previous four years' sales times the R&E ratio. As a legitimate business expense, research can already be deducted from profits. With the new credit, however, a company could not only reduce its taxable income with additional research, but actually get a 20-cent refund every year for each research dollar spent above some fixed level.

While the acceptance of a fixed base greatly brightened the picture, one problem remained. Congress insisted on authorizing the credit from year to year, essentially retaining the right to cancel it on whim. Such a climate of uncertainty "keeps industry in a constant guessing game", says Daniel Burton, executive vice president of the Washington-based Council on Competitiveness. "Companies can't assume [the tax break] in their planning budgets." Rather than planning next year's research and development on a tax credit, "they're doing R&D in spite of it", he says.

Burton, like other promoters of the R&E tax credit, is lobbying for a longer trial period. "Make it permanent for ten years and then come back and take a look at it", he suggests. But Congress, facing a huge deficit, has so far been unwilling to commit itself to losing over \$15,000 million in revenue on the basis of rosy predictions of future gain. Underlying disagreements over the credits' effectiveness are continuing questions over government sponsorship of industrial research altogether. "In this budget atmosphere, where we're asking tough questions about spending, it's unclear to me why we should exempt R&E", says Joseph Cordes, deputy assistant director for tax analysis at the Congressional Budget Office. "We've always assumed that you can't get enough R&E. But is that true? If we encourage companies to do more research than they would do otherwise, will that research be top grade? Maybe not."

Given a lack of convincing evidence of effectiveness, budget negotiators are expected once again to authorize the

credits for one year only. Although the credits are no doubt improved by last year's changes, "I wouldn't like us to simply declare victory and go home", says Cordes. "Now that we have a well-designed credit we finally have a fair experiment. We still need to watch it run." Although the new fixed-base provision will almost certainly perform better than their predecessors, evidence of its effectiveness will not be available quickly. Backers of the credits are hoping that Congress is willing to wait.

Christopher Anderson

* The Incentive Effects of the New R&E Tax Credit. Brookings Institution, July 1990.

PERPETUAL MOTION

No end in sight

Helsinki

A machine purported to be capable of perpetual motion was unveiled last week at the Heureka, the Finnish Science Centre near Helsinki. The machine (see below) consists of an embellished bicycle wheel, which spins ceaselessly, and two pendulums whose bobs move in an elliptical

Heureka

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

David Jones, on the go.

path. Static elements include a manifold of shining brass pipe, some objects resembling magnets and a grey box of unknown function.

A delegation from the Soviet Politburo which attended the unveiling were particularly astounded by the machine, saying that "Soviet experts must learn of this [discovery]". But the machine's constructor, David Jones, a guest staff member in the physical chemistry department at the University of Newcastle upon Tyne, United Kingdom, confesses that the machine is a fake. "I know the laws of physics. But by the time they have puzzled out how it works, I shall be far away," he said. Jones left Helsinki early the next day, after confirming rumours that he was linked to the DREADCO organization (see page 622).

Tytti Sutela

Fog over London

London

THE trustees of the Science Museum in London were convicted last week for "not maintaining cleaning and water treatment of their air-conditioning system in such a way that members of the public were not exposed to a risk to their health and safety", following the discovery of *Legionella* bacteria in the museum's air conditioning system. The trustees were fined £500 and ordered to pay £35,000 legal costs. The museum is considering an appeal.

This news could hardly have come at a worse time for London's research institutions and museums, currently facing financial problems.

London Zoo, showpiece of the Zoological Society of London, could close if last year's £5 million deficit is not reduced, warned Lord Peyton of Yeovil, the society's treasurer, on the release of its latest annual report. The zoo has been a drain on the society's funds for 25 years. Last year's deficit has been alleviated, to some extent, by £1.6 million of interest on a £10 million once-only government package to rescue the society in 1988, and £1.3 million of research funds from the Department of Education and Science.

At the Royal Botanic Gardens, Kew, shrinking public funds will mean that new conservation initiatives will have to rely on private money, according to Kew's director, Professor Ghilleen Prance. Kew's annual core funding of about £12 million comes from the Ministry of Agriculture, but the annual increment of 2.5 per cent is a cut in real terms. A new visitor centre and a molecular systematics laboratory were both supplied from the public purse. Prance emphasizes Kew's "positive" approach to financial constraint: a development fund to underwrite new projects has been established, and a "Friends of Kew" scheme, launched last July, now has 1,000 members.

Henry Gee

UK RESEARCH COUNCILS

New AFRC secretary

London

PROFESSOR Tom Blundell is to become secretary of the Agricultural and Food Research Council (AFRC) for five years from 1 January 1991. Blundell is a molecular biologist who heads the crystallography department at Birkbeck College London and is a member of the government's Advisory Council on Science and Technology. Throughout the 1980s, Blundell was closely involved with the UK research councils, holding posts within AFRC, the Science and Engineering Research Council and the Medical Research Council. He replaces Professor William Stewart, who became the government's chief scientific adviser on 1 October.

Peter Aldous

US "selling its children"?

Washington

CONGRESS and the US administration clashed once again last week over Japanese investment in US high technology, but while many of the issues are familiar, ominous signs of an economic downturn have raised both stakes and tempers in this ritual of Washington politics.

Last week's battle was over the sale of Semi-Gas Systems, an affiliate of Sematech, the industrial consortium that the US government helped to form in 1987 expressly to counter Japanese inroads in electronic high technology. Thanks to its association with Sematech, Semi-Gas has grown in value nearly five times and is a leader in the special high-purity gases used in the manufacture of semiconductors.

Nippon Sanso, a Japanese chemical company, found such performance hard to resist and made a bid for the company earlier this year that was approved by the White House, infuriating some in Congress (see *Nature* 346, 500; 9 August 1990). "Allowing the sale of a company that was specifically nurtured to oppose foreign competition is absurd", claimed Senator Al Gore (Democrat, Tennessee) at a hearing last week. Gore likened such actions to "selling our children". Underlying the debate is a fundamental division in US politics over foreign investment in US industry. Many Republican policymakers, led by President Bush, believe that if a foreign, even Japanese, company pays a fair price for US technology, the US economy is no worse for it. "Are we better off seeing US companies going belly-up, or are we better off seeing them foreign-owned, but on our soil, subject to our laws?" asked Stephen Canner, director of the US Treasury's office of international investment.

On the other side are those critics, such as Gore, who argue that the "fair value" of a US company goes beyond the immediate worth of its technology to the less tangible place it holds in the US economy as a whole, by virtue of the jobs it provides, the taxes it pays, its relationship with other US companies, and the cost of replacing its services, if need be. US owners, when they sell their companies to Japanese investors, are undervaluing them by not taking into consideration their full benefits to society, Gore argues. Only the federal government can have the big-picture perspective to know what a high-technology company's real worth might be, and to block the sale if it is harmful to the US economy.

Of course, Gore's argument makes most sense if there is evidence that Japanese investors are buying US companies, stripping them of their technology and shutting them down, raising the prices of their products, or restricting sales

so as not to compete with Japanese companies. Few studies have been undertaken on that subject, however. Largely anecdotal evidence compiled earlier this year by the Defense Science Board Task Force on Foreign Ownership and Control of US Industry suggests that in some cases Japanese purchase of US high-technology companies have resulted in shut-downs or created virtual Japanese monopolies that threaten US access to the technology.

On the other hand, a 1988 Department of Commerce report found that when foreign companies buy US businesses, it is largely to take advantage of the US workforce and markets, rather than to spite them. And as employers, foreign companies are often better than the US businesses they replace, the study found. The average US employee of a foreign-owned company made nearly \$39,000 per year in 1986, compared to about \$29,000 for the average employee of a US-owned company. "I think the majority of foreign companies are trying to take advantage of US business opportunities, rather than to rape the US firms and take their technology home", says Robert Lawrence, an analyst for the Washington-based Brookings Institution think-tank.

US officials argue that existing laws (the 1988 Exxon-Florio amendment in particular) allow them to halt sales only when they endanger "national security", a term they have defined as specifically relating to defence applications, not general economic well-being. "It's clear that the Exxon-Florio provision was not intended as a tool of industrial policy", said John Niehuss, senior deputy assistant secretary of the Treasury. Even if it were, he said, such interference would be unwarranted.

With no strong reason to believe that Japanese investment is undermining the US economy, Lawrence believes the only legitimate concern is the fear of foreign-controlled monopolies. Gore, at his hearing, pointed to two such cases now before US officials. One, the Connecticut-based Barden Corporation, which is the target of a takeover bid by a German industrial conglomerate, is the only source of high-precision ball-bearing that are required to keep US nuclear submarines quiet. The other, the Moore Machine Tool Company, which makes parts for nuclear weapons, is considering a bid by Fanuc, Inc., a Japanese company.

Warning the US officials that he would be watching their decision on these proposed sales, Gore offered his own guess at the reasons behind the US leniency on foreign takeovers. "I'm wondering if the United States has borrowed so much money from the Japanese, that we've become afraid to say 'no'".

Christopher Anderson

Working for the West?

London

NPO Vector, a large biotechnology research and production complex in Novosibirsk, linked to the Siberian Academy of Sciences, will sign its first business deal with the West within the next month in an agreement to sell its products through a British company, Oxford Virology. The agreement signals the arrival in Siberia of a trend that began three years ago in Moscow when the Institute of Molecular Genetics attempted to forge the first Soviet links with foreign biotechnology companies.

That institute and the M.M. Shemyakin Institute of Bioorganic Chemistry in Moscow are the leaders in the field, but very few of the discussions that Soviet researchers have had with foreign companies have led to productive collaboration. Research and development projects that require Western companies to operate within the Soviet Union have been the least successful, as laws for the operation of foreign businesses are still not fully developed.

The Siberian collaboration may do better as it involves only the marketing of Soviet products. But even this kind of agreement may run into difficulties, according to Professor Valery Soyfer, a Soviet molecular biologist now at George Mason University, Virginia. Although such agreements are very important for Soviet biotechnology, says Soyfer, with the potential to bring in badly needed hard currency, the Soviet economy is "so deeply ill" that Soviet partners may be unable to guarantee the supply of products.

John Meers, chief executive of Enzyma-

tix, a small Cambridge-based company, says there are also problems with the quality of some Soviet products. Enzymatix and some other British companies have taken a different route and licensed Soviet-developed technology to manufacture biological products in the United Kingdom. Only a few companies have so far chosen to set up joint research and development ventures with Soviet institutes. One of the largest such projects is a research laboratory and production plant at the Shemyakin Institute, set up last year by the US pharmaceutical company Monsanto and MNTK Biogen, a technology transfer organization established in 1986 to promote the commercial application of Soviet biotechnology research. The low costs of Soviet research is the principal

attraction to big Western companies. Another motivation is the desire to gain a foothold in the Soviet market, according to Rod Greenshields, managing director of GB Biotechnology.

In other disciplines, there are worries that increasing links with Western companies may divert Soviet scientists from innovative work. But commercial collaboration with the West in biotechnology is at such an early stage, says Professor Maxim Frank-Kamenetskii, from the Institute of Molecular Genetics in Moscow, that increased links can have only a positive effect on Soviet biotechnology in the short term.

The most pressing problem for academy institutes, he says, is the 'brain drain' of young scientists to the West, which may be slowed by an influx of Western finance.

Peter Aldhous

MEDICAL SCIENCE

Psychiatric abuses revealed

Munich

IN a case that has cast a shadow over the psychiatric profession in the unified Germany, the East German Volkskammer (parliament) has recommended criminal charges against psychiatrists working in the East German psychiatric clinic at Waldheim (near Dresden) who performed castrations and brain operations on at least a dozen patients against their will.

Although only three physicians were implicated directly, a special committee of the Volkskammer reported that state-employed psychiatrists across East Germany abused their professional status by revealing data about their patients to the Stasi (secret police). Stasi also issued orders to psychiatrists to detain healthy people in mental institutions in an attempt to prevent public protests, the committee reported.

The Volkskammer also recommended that the all-German parliament (Bundestag) should continue an investigation into abuses by East German psychiatrists at Waldheim and elsewhere. The case has psychiatric professionals in both parts of Germany worried about the medical credentials of their East German colleagues. The credentials of physicians trained in East Germany are generally accepted without question in the western part of Germany.

The Volkskammer considered the committee report important enough to squeeze it into the legislative calendar after midnight on the last day of East Germany's separate existence before German unification on 3 October. Extracts from the report, which has not yet been released to the public, were published in the West German newspaper *Frankfurter Allgemeine Zeitung*.

In the twelve most serious cases discovered so far, patients were castrated using X-rays or lobotomized or had parts of their brains heated, killing large numbers of cells. Other patients were reported to have been "quieted" through excessive exercise or through prolonged incarceration in small, unlit cells.

The Waldheim clinic, which unlike other East German psychiatric hospitals was attached to a prison, has been described as a kind of alternative jail for difficult inmates. Other psychiatric hospitals, such as one in nearby Hochweitschen, are also thought to have acted as a jail for citizens who were expected to become "agitators" at state-arranged demonstrations. A Waldheim city councilman was temporarily committed to Hochweitschen in 1983 as a punishment for trying to correspond with US President Ronald Reagan.

In order to find out if there were more widespread abuses, psychiatrists in both parts of Germany are urging the German Health Ministry to sponsor an investigation into conditions in psychiatric hospitals in the eastern part of Germany.

A delegation urging an investigation led by the Bonn-based 'Aktion Psychisch Kranke' (APK or Action for the Mentally Ill) was given a favourable reception earlier this year by West German Health Minister Ursula Lehr. But Manfred Bauer, an APK executive committee member, says that the Bonn government has since shed responsibility for any investigation on the grounds that the *Länder* (states), not Bonn, are responsible for an investigation. It is too early to say whether the Bundestag itself will now investigate Waldheim as a result of the Volkskammer report. Steven Dickman

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

APDEA

One of NPO Vector's more unusual products is fetal reindeer serum. Free-ranging Siberian reindeer have never been exposed to antibiotics and so can provide a culture medium that has none of the antibiotic contamination problems of the commonly-used fetal calf serum. But persuading Western scientists to risk their valuable cell cultures in an unfamiliar medium will be difficult. More realistic money-earners will be NPO Vector's range of virus insecticides and antiviral agents.

In defence of the MRC

SIR—In view of Professor Roy Calne's letter (*Nature* 347, 418; 1990), I must make it absolutely clear that the decision to discontinue funding for the MRC Medical Cryobiology Group was a peer review judgement. It was the council's Cell Board that reached the view that the work reported and proposed fell short of the current competitive standard and which decided to recommend that the group should be disbanded. This recommendation was subsequently endorsed by the council itself. As I explained in my earlier letter (*Nature* 347, 116; 1990), at each stage of the peer review process a wider scientific perspective is introduced and research boards have always had to take account of how the quality of work in one field compares with that in another.

If we were to avoid taking these hard decisions by, as Calne suggests, "spreading the hardship evenly", the end result would be that no research team was properly funded and questions of priority were never addressed. I do not believe that this would be in the interests of UK science.

D. A. REES
(Secretary)

Medical Research Council,
20 Park Crescent,
London W1N 4AL

SIR—P. J. Barnard's letter (*Nature* 347, 324; 1990) is so misleading that I feel obliged to respond.

In February 1989, the Secretary of State for Education and Science agreed increases of the order of £20 million a year in the Medical Research Council (MRC)'s parliamentary grant-in-aid for the years 1989–90, 1990–91 and 1991–92. At the same time, he agreed to provide funds for an extension of the council's AIDS Directed Programme beyond the three years originally approved. The £20 million a year was an increase of about 12 per cent — a welcome figure, but not the 53 per cent quoted by Barnard. Most of this additional £20 million was earmarked for specific new developments, such as Interdisciplinary Research Centres and the Human Genome Mapping Project.

In the financial year 1990–91, we have received a further £3.3 million which is not earmarked. But, with pay inflation running at some 9 per cent, the council is having to find about £11 million a year to meet the additional costs of paying its staff and of salaries paid under grants.

Barnard also refers to the council's decision to close the Molecular Neurobiology Unit on her husband's retirement as director. She does not refer to the council's decision, at the same meeting to see whether a leader of high scientific standing could be recruited to lead a new pro-

gramme in the field of molecular neurobiology. If that were possible, the precise organizational relationships of the new team within the MRC would be a matter for negotiation and dependent on requirements for scientific support and collaboration. The council endorsed the Neurosciences Board's hope that there would be opportunities for working in close association with teams in the Laboratory of Molecular Biology.

These longer-term plans for supporting the field of molecular neurobiology are in no way affected by the decision to withdraw resources early from the existing unit in order to help fund some high quality programme and project grant applications that would otherwise have been declined. The savings are being targeted on those sections of the unit that were judged as less competitive following peer review.

NICK WINTERTON

(Head of Secretariat)

Medical Research Council,
20 Park Crescent,
London W1N 4AL, UK

BSE susceptibility

SIR—Our paper on the inheritance of susceptibility to bovine spongiform encephalopathy (BSE) in pedigree Holstein Friesian cattle (*Veterinary Record* 126, 5–8; 1990) is relevant to the question raised by your correspondent O. H. G. Sparrow (*Nature* 346, 694; 1990) as to whether BSE is gene-linked.

Our findings relate to the six-generation pedigrees of 14 BSE-affected cows and 14 cows drawn at random from a multiple-case herd and also compared the number of sires and maternal grandsires of the affected cows with 200 samples of 14 randomly drawn unaffected cows per sample from the same herd. The six-generation pedigree showed a similar pattern of occurrence of five bulls and their combinations in both BSE-affected and control groups. The nine sires and eight maternal grandsires of the BSE-affected cows were comparable with the modal number of nine sires and nine maternal grandsires in the 200 samples and were therefore as would be expected solely on the basis of the breeding structure of the herd. Forty-four (73 per cent) of the 60 BSE-affected cows were familial in origin with first or second degree relatives with the disease. A study of family pedigrees over three or four generations of 43 BSE-affected cows and one BSE-affected bull showed that one cow and 11 bulls were the heads of families. In the largest family group, with nine BSE cases, the head of the family was an imported Holstein cow.

These heads of families were all of

Canadian Holstein or Dutch Friesian heredity. The fact that no single common ancestor could be incriminated in the Holstein Friesian breed as transmitting BSE, that BSE was diagnosed in other breeds and crossbred animals within a relatively short period after it occurred in the Holstein Friesian and that similar encephalopathies have occurred in other related species such as antelopes within this period suggests that susceptibility to the disease and not the disease itself is inherited.

In a study supported by the Ministry of Agriculture, Fisheries and Food (MAFF) we are now analysing in more detail an extensive set of records of pedigree herds provided by the Holstein Friesian Society of Great Britain and Ireland (HFS). Using information provided by MAFF Central Veterinary Laboratory, Weybridge, on confirmed cases and feeding regimes, we are investigating the extent to which the susceptibility of cattle to BSE may be inherited.

W. V. S. WIJERATNE

44A Berkeley Avenue,
Reading,
Berkshire RG1 6JE, UK

R. N. CURNOW

Department of Applied Statistics,
University of Reading,
Whiteknights,
Reading,
Berkshire RG6 2AN, UK

Hunger strike

SIR—Your readers should know that Victor Ambartzumian, one of the most distinguished astronomers of this century, the former president of the International Council of Scientific Unions and International Astronomical Union, and present president of the Academy of Sciences of Armenia, has begun a political hunger strike in Moscow. The 82-year-old scientist has thus joined the well-known writer Zori Balayan whose hunger strike began on 10 September. The two members of the Soviet Supreme Council are protesting against the violation of human rights in the mountainous Karabagh Armenian region of Azerbaijan. For the third year, this region is under total economic, communications and information blockade, with the result that the problem is widely represented as an ethnic or even religious problem. But, as has often been stressed by the late Andrei Sakharov, the problem of Karabagh is a problem of elementary human rights and discrimination against Armenians.

The time has come for the international scientific community to protest strongly at these terrible circumstances.

V. GURZADYAN

Yerevan Physics Institute,
Yerevan 375036,
Armenia, USSR

Hands up for toxic waste

Herbert Inhaber

Proposals for the siting of toxic waste plants often meet local opposition. This hurdle could be overcome if local governments were to place 'compensation' bids for burying waste, while maintaining environmental standards.

THE authorities in many countries are becoming used to vehement local objections when proposals to bury toxic and nuclear wastes are floated. Usually the protests are confined to words, but occasionally they descend to violence. In the United Kingdom, for example, police investigated a case of arson at a proposed nuclear waste site in Fulbeck, Lincolnshire, in 1987; and the Scottish National Liberation Army claimed responsibility for a fire at Glensanda, Scotland, in 1989 because the location was thought to be a candidate for a nuclear-waste facility.

In the United States, the first nuclear-waste riot occurred in April in Allegany County, New York. As officials attempted to inspect the site of a proposed low-level radioactive waste facility, protesters wearing masks charged police lines on horseback, rolled large snowballs down hillsides and erected roadblocks¹. Six people handcuffed themselves to each other and formed a human barricade across a bridge.

Can a system be devised to site hazardous and nuclear wastes — and a host of other undesirable facilities — without such confrontation? The reverse Dutch auction is such a process. It requires no coercion while preserving environmental and safety standards. The principle is to encourage a local community to volunteer to accept the waste site.

According to an opinion poll, the US public's two greatest environmental concerns are not global climate change, habitat alteration or other macro-environmental issues. Rather, they are active and abandoned hazardous waste sites, with about two-thirds of the populace being troubled by their real or perceived risks².

Those who protest against the siting of 'locally unwanted land uses' are derided as unreasonable by the authorities. But as Bingham writes³, "Although it may sound heretical or obvious, *local residents are acting rationally in opposing hazardous waste facilities* . . . The reason local residents oppose new facilities is that they have every incentive to do so — the new facility makes them worse off. Thus, the most direct way to respond to such opposition is to change the incentives that motivate human behavior".

Compensation schemes have already been proposed for some of these unwanted sites. In New York State, for example, the siting commission will make

payments of about \$1 million a year for a planned low-level facility⁴. The Federal Nuclear Waste Policy Act, as amended in 1987, would pay between \$10 and \$20 million annually to a state that accepted high-level nuclear wastes. But schemes such as these have a major defect. The amounts are set by a central authority, with little or no local input. As one of the main concerns of those affected is a lack of control over the process, compensation schemes simply accentuate their impotence.

The reverse Dutch auction differs from these schemes in that it allows people living close to a potential waste site to set their own level of compensation, empowering them to accept or reject the proposal. The auction scheme is based on simple principles, already applied to a wide range

of human behaviours. Any general would say that an army of volunteers will fight more effectively than a group of sullen draftees. So the reverse Dutch auction is intended to generate a volunteer community.

A standard, or English, auction, has a rising price and more than one bid. But we do not want more than one willing region. A Dutch auction has a falling price — sometimes set on a clock — and only one bid. When the bidder signals, the auction is over. English and Dutch auctions deal with desirable objects. In the case of unwanted and undesirable land uses, the auction must run in *reverse*.

The first step in the reverse Dutch auction is the announcement of the environmental or risk ground rules, including a maximum population density,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Greenpeace and other organizations reflect the strong public opposition to some of the present arrangements made to dispose of toxic waste. In this case, the controversy was over the proposal to dump UK nuclear waste in the Atlantic.

Greenpeace/Greig

groundwater considerations, geological regulations, endangered species rules and so on. Because no specific local region is identified at the beginning of the process, these ground rules can be chosen relatively objectively. In the United States at present, ground rules are sometimes slanted to ensure that a specific area or region is chosen. For example, when the Department of Energy developed mathematical criteria for its proposed monitored retrievable storage facility for high-level nuclear wastes, it set up the formulae so that Oak Ridge, Tennessee, was virtually the only site that qualified⁵.

To make a bid from a local region for a site as inexpensive as possible, the siting authority will pay for the hiring by counties of consultants to help find a suitable site consistent with the ground rules. But local regions will not be able to hire a consultant to demonstrate that it is not suitable. The state of Nevada, for example, has spent a large proportion of the money allocated to it by the federal government in trying to show that it is not appropriate for high-level wastes. Under the new scheme, there would be no reason to hold these pseudo-scientific battles: a region that did not want to host a site would simply not bid for it.

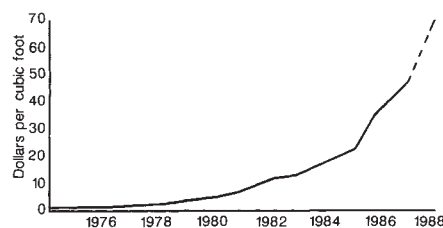
When the auction begins, the siting authority publicly offers any volunteer region an amount of money if it can find an environmentally acceptable site. The bonus keeps rising steadily, to entice regions that had hitherto sat on the sidelines to select a site. Eventually, one region will suggest a site when the true social cost of the facility — the price the rest of us have to pay to have the site away from *our* back yard — is reached.

The system is procedurally analogous to that used by the US Nuclear Regulatory Commission, its predecessor the Atomic Energy Commission and other regulatory authorities around the world. The NRC has never chosen a reactor site; this was always done by a utility, which presumably had a reason for building the facility. The NRC approved or disapproved the site only after submission by the utility.

Evaluation

After a bid has been received, the site is evaluated on the basis of the prearranged ground rules. As a volunteer region would not bid unless it were confident of succeeding, chances of disapproval are small. Nonetheless, the siting authority has the right to reject the bid on environmental grounds. In that case, the auction recommences.

The volunteer county will, by definition, have a consensus in favour of the bid. If it did not, it would not bid. Elected officials will be confident they will not be defeated at the next poll on this basis. Gone will be their slogan "not in my term of office". A large proportion of the bonus



Charges for low-level radioactive waste burial at facilities in Barnwell, South Carolina⁷. Hatched line, projected costs.

is paid into a trust fund until the scientific evaluation is completed. Democratic legislatures often undergo a tortuous and meandering process to appropriate money. In the new scheme, a lengthy delay in allocating the bonus means that the consensus built up in the community in favour of the site will evaporate. If the site is disapproved, the bonus of course reverts to the siting authority. If the site is approved, the rest of the bonus is paid to the volunteer region for any local purpose — reducing taxes, building roads, improving services, building a trust fund for future generations, health or radiological monitoring, or any combination of them.

Are the incentives offered by the reverse Dutch auction a form of bribe? I think not. First, in any system, finance and safety and health considerations are to some extent intermingled. But, more specifically, the reverse Dutch auction does not have three of the main elements of a bribe. All steps will be public, whereas a bribe is given under the table. Second, bribes are targeted, that is, they are made with a specific person or agency in mind. In the new scheme, the volunteer area is self-selected, so there are no targets. Finally, a bribe involves an illegal activity. Almost everyone is in favour of burying toxic and radioactive wastes, so long as it is not in their back yard. The disposition of these wastes, under appropriate environmental regulations, is clearly encouraged and legal. The reverse Dutch auction is an example of the use of an incentive, rather than a bribe, to solve a difficult problem.

It is possible that neighbouring regions would object to a nearby region volunteering a site, especially if the site were to discharge effluent into a watercourse flowing into them. Almost all analyses of risk predict that there would be little or no effluent. Nevertheless, a consortium of regions encapsulating the area thought to be at risk might bid together within the auction framework, dividing the bonus according to anticipated risk.

Will the new scheme cost too much? No auction has a final price fixed in advance. As a result, policymakers may be concerned about unlimited liability. But this problem can be solved here by setting a cap on the bonus to be offered. That cap may have to be substantial — a waste-disposal company has offered the residents of a town in western New York

about \$960 per person annually for the right to build a landfill for non-hazardous wastes⁶. Most people would regard hazardous wastes as being psychologically more damaging than their non-hazardous counterparts, and therefore require higher compensation for that damage.

Costs for waste disposition have risen dramatically in any event. The graph shows the price charged per cubic foot for the burial of low-level radioactive waste at a US disposal site⁷. It depicts an approximately 50-fold increase over slightly more than a decade, without the waste itself becoming any more dangerous. In New York State, about \$37 million has been spent to find a site for low-level radioactive waste, without an ounce yet being buried⁸. In the new scheme, no money would change hands until a specific site was approved and agreed upon. There is a much closer relationship between the product — waste disposition — and financial outlays than under the present system.

Incomprehension

Economists and risk analysts have been urging a more sensible waste siting system for some time. While it would take too much space to list all the contributions, Kunreuther⁹ and O'Hare¹⁰ have noted the excessive complication, with few results, of the present system. For example, the low-level radioactive waste bill in Massachusetts was 49 pages long, in an effort to be fair to all conceivable parties. It was admitted on the floor of Congress that none of its members could understand the amendments made in 1985 to the law governing this type of waste disposal¹¹. Both Kunreuther and O'Hare have urged the use of incentives to site these wastes.

The reverse Dutch auction allows a market-based solution to a seemingly insoluble problem, at the same time protecting environmental standards. There are few, if any, extant proposals for siting toxic and nuclear wastes combining these features. Because it is market- rather than bureaucracy-driven, it removes the alleged coercion that has marked most attempts to site these facilities. □

Herbert Inhaber is at Ecology and Environment, Inc., Buffalo Corporate Center, 368 Pleasantview Drive, Lancaster, New York 14086, USA.

1. Simon, P. *Buffalo News* **A1** (6 April, 1990).
2. Roberts, L. *Science* **249**, 616–618 (1990).
3. Bingham, G. & Miller, D.S. *Natural Resources Lawyer* **17**, 473 (1984).
4. *Buffalo News* **B2**, (13 September 1989).
5. Inhaber, H. *J. haz. Mat.* **15**, 241–263 (1987).
6. *Buffalo News* **B5** (8 August, 1990).
7. Shaw, R. A. in *Low-Level Radioactive Waste Regulation: Science, Politics and Fear* (ed. Burns, M.E.) 119–140 (Lewis, Chelsea, Michigan, 1988).
8. Inhaber, H. *Empire State Report* **17**, 41–42 (1990).
9. Kunreuther, H. & Kleindorfer, P. *Am. Econ. Rev.* **76**, 295–299 (1988).
10. O'Hare, M. *Public Policy* **25**, 407–458 (1977).
11. Colglazier, E.W. & English, M.E. in *Low-Level Radioactive Waste Regulation: Science, Politics and Fear* (ed. Burns, M.E.) 215–238 (Lewis, Chelsea, Michigan, 1988).

How to shadow noisy chaos

Chaotic data in the strict sense cannot usually be demonstrated to be such by computer simulation, if only because of the unavoidable rounding errors of computation. But now there may be a solution of the difficulty.

THE discipline of chaos (not a contradiction in terms, but the name of the academic pursuit) may be most of all beguiling because everybody can play. Given a programmable hand calculator and an interesting idea for a way of mapping the set of real or even complex numbers onto itself, a chaotic sequence of numbers can be plucked from the air. Given a personal computer, particularly one with a colour monitor, the opportunities for fun and games are almost indefinitely expanded. This seems to be the spirit in which some eventually scholarly studies have begun.

But in the real world, this is not how people are most often brought face to face with chaos. Instead, there will be a sequence of numbers gathered by measurement or observation, perhaps representing the properties of some system at successive intervals of time. The task is to judge whether the underlying behaviour is chaotic in the strict sense or whether, alternatively, the appearance of chaos is a consequence of the errors of measurement or the finiteness of the sequence of observations. Many of the recent suggestions that the motion of asteroids or minor planets is chaotic motion are probably, in this respect, still undecided (and may even be undecidable on the data so far available).

So it is natural that people should be seeking ways in which chaos can be recognized unambiguously. The ideal, given a set of sequential data, might be thought to be a successful guess at the underlying law relating one member of the sequence to its predecessors and a simulation of the data-set by an appropriate computer calculation. But if the data-set is indeed chaotic, such an ambition will be intrinsically unattainable, simply because the rounding errors arising in computations with numbers whose precision is necessarily finite will indefinitely diverge with successive iterations.

That point is put vividly by Celso Grebogi, Stephen M. Hammel, James A. Yorke and Tim Sauer, who point out that computation with 8-figure (decimal) precision of a much studied set of chaotic trajectories will, after only 20 iterations, magnify errors of one unit in the last decimal place until they are comparable with the numbers which are the objectives of the exercise.

This is nothing but the familiar complaint of numerical weather forecasters that the range of their forecasts, similarly based on iterative procedures, is limited

by the way in which the atmospheric consequences of the flapping of a single seagull's wings may ultimately become a complete weather system. The ostensible starting conditions, the observed state of the weather at the beginning of the forecasting period, have been lost and buried in the emergent chaos. Plainly there is, in general, no hope of checking that a sequence of data represents chaotic behaviour by guessing an underlying rule and checking that it reproduces observation; however good the guess, computational rounding errors would eventually fail to reproduce the data accurately.

So what can be done instead? Grebogi and his colleagues (Grebogi and Yorke are from the University of Maryland at College Park, Hammel is from the nearby Naval Surface Warfare Center and Sauer from the George Mason University, the other side of the Potomac) set out to define the conditions in which a calculated data-set may be held to shadow a chaotic data-set, perhaps with an adjustment of the initial conditions (*Phys. Rev. Lett.* **65**, 1527; 1990).

Shadowing is nothing but the process in which espionage agents are supposed perpetually to be engaged — staying close to some other person. The equivalent of that, in chaos, is telling whether a set of noisy data, one generated by computer iteration with rounding errors, can be followed within predefined limits by some 'true' iteration. The conclusion is that shadowing is indeed possible in most cases, but the proof is at once intricate and interesting.

By way of illustration, Grebogi and his associates use the case (among others) of the harmonically driven pendulum, described by the nonlinear equation $\ddot{y} + \sin y = f_0 \cos t$, where the *umlaut* represents double differentiation of the angular displacement y with respect to time. (The left-hand side is the textbook problem of a pendulum with simple harmonic motion except that the simplification to small displacements is avoided.) They first generate a set of data for a pendulum starting literally at rest, using 60 significant data-bits in their representation of y . They integrate the equation numerically, with a time-step of $\pi/1,000$, and let the computer run for 10 million iterations. This is the 'noisy' data-set, in which (with 60 data-bits) the rounding errors must be less than 10^{-18} .

How can these data be shadowed? Conceptually, Grebogi and his colleagues use figurative parallelograms supposed to

contain a small portion of the true graph of $\ddot{y}$ against y , the quantities calculated in the numerical program. The trick is then to iterate both the curve and the parallelogram by means of the nonlinear equation for the driven pendulum. Briefly, it emerges that it is always possible to construct a true curve shadowing the empirical data provided that two conditions are satisfied — that successive parallelograms overlap each other in the sense that if they stretch in one direction they get narrower in the other, and that the angle they contain never shrinks to zero.

But is it likely that these conditions will be fulfilled? To increase the chances, Grebogi and his colleagues use a numerical technique for smoothing out the noise in data without altogether getting rid of its chaotic character. For practical purposes, this boils down self-consistently to correcting each pair of noisy data (for $\ddot{y}$ and y) by amounts related to the sensitivity of the two variables to the underlying equation of the driven pendulum.

That the procedure can be followed without getting into intellectual knots has been verified by using 96-bit arithmetic (corresponding to a precision of 10^{-28}). As a smoothing technique, the proposed drill seems to be superbly efficient, with the number of significant digits apparently doubling at each step. But its efficiency requires foreknowledge of the underlying laws, which may not be the case for sets of empirical data.

And sometimes it may not work well at all. While it seems always to be possible to devise a true chaotic trajectory that will shadow noisy data almost everywhere, there will be points in the data-set (called 'glitches') at which the recommended recipe will not function and at which even the smoothing procedure may be ineffectual. So the practical question becomes that of telling how close will be the shadowing of the empirical data by the 'true' chaotic trajectory, and how frequently the process will be interrupted by a glitch.

Grebogi *et al.* have no formal proof to offer, but their arithmetic is persuasive. And they offer the conjecture that for data with noise δ (the root-mean-square of the rounding errors, say), the true curve constructed by their recipe will differ from the data by less than $\sqrt{\delta}$ and they guess that the shadowing will persist for $1/\sqrt{\delta}$ iterations until it hits a glitch. That, at least, is something into which people with Crays will be able to sink their teeth.

John Maddox

Deficiencies in sight with the candidate gene approach

Thaddeus P. Dryja

PHENOTYPES affecting the vertebrate eye are governed by a large number of genes, reflecting the complex biochemistry and cellular physiology necessary for vision. Mutations of these genes, many of which are expressed only in the eye, generally will not hinder the viability of individuals who carry them. The result, in humans, is a panoply of phenotypes ranging in severity from inconsequential (blue eyes or brown) to devastating (congenital blindness). This collection also includes genetic diseases that primarily result in the degeneration of the retina. Our understanding of them has been rudimentary and, until recently, biochemical defects were known for only a few rare forms of hereditary retinal degeneration in humans, such as Refsum's disease, abetalipoproteinaemia and gyrate atrophy. But now, they are yielding to the onslaught of molecular genetics technology. The most recent conquests are choroideraemia in humans, for which a candidate gene is reported by Cremers *et al.* on page 674 of this issue¹, and hereditary retinal degeneration in *rd* mutant mice, for which the precise biochemical defect is described by Bowes *et al.*² on page 677.

Gene isolation

Choroideraemia is a sex-linked form of hereditary retinal degeneration, the disease locus being on the long arm of the human X chromosome (Xq21). Males with the disease gradually lose vision in the mid-periphery of the retina, with a progressively decreasing island of central vision that is extinguished usually by the age of 60. The candidate gene was isolated using what has become the 'classic' approach to the cloning of a human disease gene based only on its chromosomal assignment. The first step is the laborious and sometimes serendipitous identification of DNA probes close (less than a few hundred kilobases) to the disease gene. A restriction map of the region surrounding the probe sequence is then constructed. Finally, one searches for messenger RNA transcripts derived from the region, usually by finding and using as a probe a sequence in the mapped region that has been conserved during evolution. Examples of human disease genes that have been cloned by variations of this approach are those for chronic granulomatous disease, Duchenne's muscular dystrophy, retinoblastoma, cystic fibrosis, DCC (for deleted in colon carcinoma), Wilms' tumour and von Recklinghausen's neurofibromatosis.

For the choroideraemia gene, the

unknown sequence DSX165 was the key probe, as it can detect deletions or translocations of Xq21 in some patients with choroideraemia. About 45 kilobases of DNA surrounding DSX165 were mapped, and 15 single-copy probes isolated; two of the probes contained highly conserved sequences, and one of those two strongly hybridized with a distinct mRNA transcript expressed in the retina. Surprisingly, the gene is also expressed in cervical carcinoma (HeLa) cells and lymphoblasts. As Cremers *et al.*¹ point out, choroideraemia might therefore be a retinal degeneration secondary to a widespread metabolic defect, as is the case for Refsum's disease, abetalipoproteinaemia and gyrate atrophy. In fact, gyrate atrophy, which is caused by a deficiency of the enzyme ornithine aminotransferase, is clinically similar to choroideraemia, and in the past some ophthalmologists felt that choroideraemia and gyrate atrophy were forms of the same disease. The truth will be revealed in the near future, when the 'choroideraemia protein' predicted by the mRNA sequence is identified.

Despite the successes of 'reverse genetics', I believe that the identification of other human genes causing retinal degeneration (and perhaps many other human diseases) will come to rely more heavily on an alternative method, the 'candidate-gene' approach. With this technique one picks a cloned gene that is known to have a role in the physiology of the diseased tissue and then searches for mutations of the gene in patients with the disease in question. Recent technological advances for detecting point mutations in a candidate gene, such as those based on single-strand conformation polymorphisms³ and direct sequencing of DNA using the polymerase chain reaction, make it possible to screen hundreds of patients, showing dozens of phenotypes, in a few months once the sequence and intron-exon structure of the gene is known. If a mutant sequence can be successfully correlated with a particular phenotype, then the known biological function of the candidate gene provides insights into the biochemical defect causing the disease. This approach conveniently bypasses the need for viable but diseased human retinas, which are available only rarely.

The candidate-gene technique is appealing to those enamoured of the scientific method, because it allows the formulation and testing of hypotheses about the aetiology of a disease on the basis of its clinical features and the role in the retina of a particular candidate protein. There

are numerous candidate genes available for studies of the hereditary retinal degenerations. For example, there are cloned genes encoding proteins involved in phototransduction (opsin, transducin, 48K protein, phosphodiesterase and so on), vitamin A metabolism (interphotoreceptor retinol binding protein, cellular retinaldehyde binding protein) and the structure of the disk membrane (peripherin). Many of these proteins consist of separately encoded subunits, such as the α , β and γ subunits of transducin. Furthermore, some of the subunits are specific for certain photoreceptors — for example, the transducin α subunit differs between rods and cones.

Retinal abnormalities

Do naturally occurring mutations of any of these genes lead to retinal abnormalities *in vivo*, as one might expect? The answer is yes, because (1) a missense mutation of the rod opsin gene occurs in some patients with autosomal dominant retinitis pigmentosa, a form of retinal degeneration⁴; (2) the gene that is mutant in the *rd*s (for retinal degeneration slow) mouse normally encodes the rod disk membrane protein peripherin⁵; and (3) Bowes *et al.*² now show that the gene encoding the β subunit of rod phosphodiesterase is responsible for autosomal recessive retinal degeneration in the *rd* mouse. Actually, before any of these discoveries in vertebrates, it was found that some *Drosophila* genes causing photoreceptor degeneration normally encode proteins involved in phototransduction or the cytoskeleton of photoreceptors⁶⁻⁸.

For human retinal diseases, the task at hand is to use existing techniques to link each of the growing number of cloned candidate genes with one or more retinal degenerations or nonpathogenic phenotypes. We should expect that in the next decade many if not all of the genes involved specifically in retinal physiology will be found to be mutant in some people with visual deficits. For the longer term, let us hope that the knowledge gained from the resulting inventory of biochemical defects will provide clues to therapies for the patients concerned. □

Thaddeus P. Dryja is in the Department of Ophthalmology, Harvard Medical School and Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114, USA.

1. Cremers, F.P.M., van de Pol, D.J.R., van Kerckhoff, L.P.M., Wieringa, B. & Ropers, H.-H. *Nature* **347**, 674–677 (1990).
2. Bowes, C., Li, T., Danciger, T.L., Baxter, L.C., Applebury, M.L. & Farber, D.B. *Nature* **347**, 677–680 (1990).
3. Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. *Genomics* **5**, 874–879 (1989).
4. Dryja, T.P. *et al.* *Nature* **343**, 364–366 (1990).
5. Connell, G., Boscom, R., McInnes, R. & Molday, R.S. *Invest. Ophthalmol. Vis. Sci.* **31** (suppl.), 309 (1990).
6. Washburn, T. & O'Tousa, J.E. *J. Biol. Chem.* **264**, 15464–15466 (1989).
7. Shieh, B.-H., Stamnes, M.A., Seavello, S., Harris, G.L. & Zuker, C.S. *Nature* **338**, 67–70 (1989).
8. Montell, C. & Rubin, G.M. *Cell* **52**, 757–772 (1988).

Reversing centrifugal forces

Bruce Allen

NEAR a massive compact body, such as a black hole, the centrifugal force reverses direction, according to new work by Abramowicz and coworkers¹⁻³. The result adds another item to the list of strange relativistic effects associated with gravitationally compact objects. In addition to its intrinsic interest⁴, the reversal of the centrifugal force provides a simple intuitive explanation for a number of previous predictions concerning the behaviour of viscous fluids near a black hole. These effects have important consequences for the accretion disks formed when matter, attracted by enormous gravitational forces, spirals towards such compact bodies.

The reversal of the centrifugal force can be described in terms of a simple experiment. Imagine attaching a rocket to a small test mass, and using the rocket to maintain the mass on a circular trajectory centred about a black hole. The rocket is set up so that its thrust is exerted perpendicular to the path of the test mass; thus the rotational velocity of the particle on its circular path does not change with time.

Far away from a black hole, or in empty space, the rocket must exert an inwards force, which must be increased if the rotational velocity of the test mass on its circular path is increased. If the rocket's motors are shut off, the test mass will fly off along a straight line, at a tangent to the circular path.

What Abramowicz and his collaborators have shown is that if the test mass is moving on a circular trajectory very close to a black hole, the rocket must exert an outwards force which increases if the rotational velocity of the test mass is increased. This is contrary to ordinary intuition; we would expect that increasing the rotational velocity of the test mass would tend to make the mass fly away from the black hole, so that a stronger inwardly-directed force would be needed to maintain a circular path. One can imagine building a track which circles around the black hole, and then riding a cart around the track. As the cart begins to

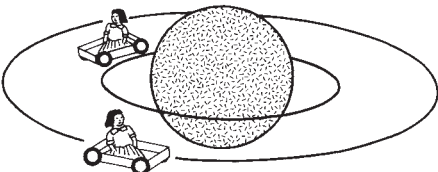


FIG. 1 A pair of circular tracks encircle a black hole, and riders in carts move around the tracks with constant velocity. The inner track lies inside a radius $3/2 r_s$, and the outer one lies outside that radius. If the riders' carts are speeded up, the outer rider is crushed more firmly against the outside of her cart; the inner rider, however, is crushed more firmly against the inside of her cart.

rotate faster and faster around the track, the rider inside finds herself crushed up more and more firmly against the inside of the cart (Fig. 1).

How close does one have to be to the black hole before the centrifugal force reverses its direction? Suppose the black hole has mass m , and we denote the speed of light by c , and Newton's gravitational constant by G . For a compact body of mass m , it is convenient to introduce the 'Schwarzschild' or 'horizon' radius $r_s = 2Gm/c^2$. For the mass of the Earth, r_s is about a centimetre; for astrophysically interesting black holes r_s would be at least a kilometre. The physical significance of this radius is that if the mass m is collapsed into a sphere of radius r smaller than r_s a black hole forms. In that case, any object which enters within the sphere $r = r_s$ is doomed to fall into the gravitational singularity at $r = 0$; it can never escape to the region outside the horizon.

The surprising effect found by Abramowicz and collaborators is that for circular trajectories in the equatorial plane, whose radius is less than $3/2 r_s$, the outwards force required to maintain the test mass on the circular trajectory increases with increasing rotational velocity. The roots of the idea can be found in earlier work by Abramowicz and Lasota^{4,5}, where they show that for a test mass circling the black hole at radius $3/2 r_s$, the force required to maintain the test mass on the path is independent of its velocity. This critical radius $3/2 r_s$, at which the centrifugal force reverses sign, happens to be the same as the radius at which a photon orbits a black hole. This is no coincidence: analysis using the optical-metric formalism of Abramowicz, Carter and Lasota⁶ shows that, in a static space-time, the centrifugal force reverses precisely when photon geodesics (in the optical three-metric) curves inside, rather than outside, the path of the test mass. Another, equivalent statement of the effect is that the centrifugal force repels trajectories away from unstable circular photon orbits, and attracts them towards stable ones.

Fortunately, this surprising result is not a question of faith — the reader with a background in general relativity can verify the main result in an afternoon. The details can be found in section 4.1 of ref. 3; indeed, within a few years this effect will probably be presented as an exercise for students of relativity. One simply calculates the acceleration of a test mass moving with constant rotational velocity about a circular path (Fig. 2). Choosing units with $G = c = 1$, one finds that the velocity-dependent part of the acceleration of the test mass is perpendicular to the surfaces

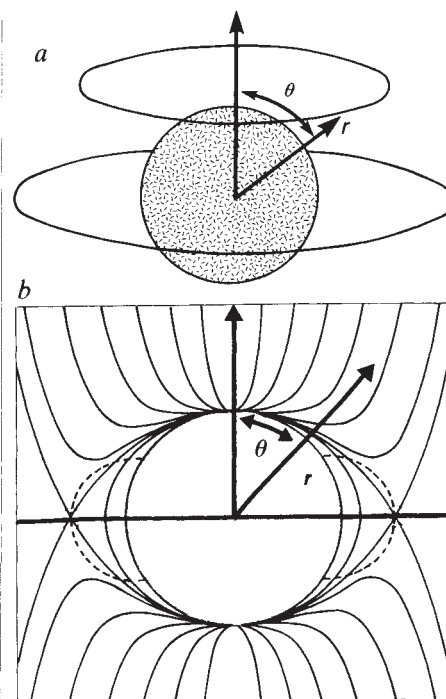


FIG. 2 Reversing centrifugal forces. Test-masses are moved in circular paths, centred around the black holes (a), which are parallel to the equatorial plane, at radius r from the centre of the black hole, and an angle θ from the azimuthal axis. b, For a particle moving in such an orbit, the centrifugal force is normal (perpendicular) to the surfaces known as von Zeipel cylinders defined by $r^2 \sin^2 \theta / (1 - 2m/r) = \text{constant}$. Within the 'rotosphere' (dashed line) the component of the centrifugal force perpendicular to the cylinders $r \sin \theta = \text{constant}$ points inwards rather than outwards.

known as von Zeipel cylinders (Fig. 2b). Inside the 'rotosphere' ($r < 3m \sin^2 \theta$) the component of the acceleration normal to the cylindrical surface ($r \sin \theta = \text{constant}$) points outwards rather than inwards. Thus, inside the rotsphere, the centrifugal force reverses sign.

There are a number of interesting implications of this result, and Abramowicz in fact claims that all dynamical effects associated with the centrifugal force change sign inside the rotsphere. These effects are expected to be most important for the behaviour of the accretion disks formed when matter is pulled into a black hole by gravitational forces. These accretion disks might contain a considerable quantity of material, and it is believed that they are formed when the gravitational field of the black hole tears matter off the surface of a

1. Abramowicz, M.A. & Prasanna, A.R. *Mon. Not. R. astr. Soc.* **245**, 720–728 (1990).
2. Abramowicz, M.A. & Miller, J.C. *Mon. Not. R. astr. Soc.* **245**, 729–732 (1990).
3. Abramowicz, M.A. *Mon. Not. R. astr. Soc.* **245**, 733–746 (1990).
4. Abramowicz, M.A. & Lasota, J.-P. *Acta Phys. Pol.* **B5**, 327–329 (1974).
5. Abramowicz, M.A. & Lasota, J.-P. *Acta Astr.* **30**, 35–39 (1980).
6. Abramowicz, M.A., Carter, B. & Lasota, J.-P. *Gen. Relativ. Gravit.* **20**, 1173–1183 (1988).
7. Anderson, M.R. & Lemos, J.P.S. *Mon. Not. R. astr. Soc.* **233**, 489–502 (1988).
8. Chandrasekhar, S. & Miller, J.C. *Mon. Not. R. astr. Soc.* **167**, 63 (1974).

companion star or stars orbiting nearby.

The importance of these accretion disks stems from the simple fact that a black hole itself is black — it does not emit any light or other radiation that would make it visible from the Earth. But it might become visible indirectly as matter in the accretion disk spirals into the black hole, undergoing large accelerations and shocks, and emitting electromagnetic waves. These effects are most important near the horizon of the black hole, where the gravitational field is strongest.

To understand the behaviour of accretion disks, one needs to know, for example, how the disk transports energy and angular momentum. The reversal of the centrifugal force provides a simple explanation for several earlier results concerning the behaviour of viscous fluids near a black hole. If one spins a dinner-plate covered with syrup (a viscous fluid) the syrup spreads outwards, transporting

the angular momentum outwards: but near a black hole, angular momentum is transported inwards in such a fluid⁷. We can now see that this is because the system attempts to minimize its energy while conserving angular momentum, reversing the direction of the viscous torque¹. Another surprising result, which may also be easily understood via the reversal of the centrifugal force, is that slowly rotating contracting ellipsoids reach a maximum eccentricity, and then become more spherical as they collapse⁸. The Rayleigh criterion for the stability of an angular momentum distribution is similarly reversed inside the rotosphere. The hope must be that the new effect will also cast new light on the astrophysical importance of black holes, their behaviour and appearance. □

Bruce Allen is in the Department of Physics, University of Wisconsin, Milwaukee, Wisconsin 53211, USA.

BEHAVIOURAL ECOLOGY

How to find the top male

Andrew Pomiankowski

LEKS are assemblies of males that females visit solely for the purpose of mating. An outstanding puzzle about leks is the extreme skew in the distribution of male matings — one male achieves nearly all of the copulations, most of the others none. Several male cues are known to affect female mate choice (for example strutting rate in the sage grouse¹ and tail length in Jackson's widowbird²); but these traits do not accurately predict the identity of the top male, nor can they alone account for the unanimity of female choice. Some other factors must be involved, and one suggestion is that females copy the mate choice of others. Wade and Pruett-Jones³ have shown how, in theory, even a weak tendency to copy could account for the skew in male mating success, and this prediction found support from studies of grouse leks reported at a recent meeting*.

In many lek species there are several opportunities for females to copy. Females often visit in groups and make non-mating visits to the lek, when other females are present, before choosing a mate. Most theoretical models of sexual selection have ignored this, and have assumed that females choose males independently.

To study copying behaviour, Wade and Pruett-Jones draw an analogy with Polya's Urn model. In Urn models coloured balls are picked at random from an Urn and replaced with several additional balls of the same colour⁴. The distribution of balls

after n selections is given by the Polya-Eggenburger equation. Copying is similar to this if the probability that a female chooses to mate with a male is dependent

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Sage grouse — strutting his stuff.

on the number of times that he has been chosen before.

This equation shows that whenever copying occurs the variance in male mating success increases: more males have no mates and a few males obtain a large number of matings. Interestingly, this occurs even when females have no previous preferences for particular males. But if males do differ in their attractiveness to females in the first place, copying will increase differences and cause a marked skew in the distribution of matings.

At the meeting, there were reports of two new studies of mating sequence on the

lek that provide evidence that females do indeed copy each others' choice. On days when large numbers of sage grouse hens visited the lek, hens were more likely to choose a male already selected by others (R. Gibson, J. Bradbury and S. Vehrencamp, University of California). Successful males had more matings than would be expected by independent choice, even when relative male attractiveness (calculated over the whole season) was controlled for — so it was not just that more attractive males had more mates. A similar pattern occurs in black grouse (J. Høglund, A. Lundberg, Uppsala University; R. Alatalo, Jyväskylä University). Males that mated more than once tended to copulate in sequence, either on the same or the following day⁵.

Copying appears to be due to females finding males with other females close-by more attractive. In both fallow deer⁶ and sage grouse (R. Gibson *et al.*) the rate at which females enter male territories correlates with the number of females already present. Further support is provided by an experiment in which stuffed female black grouse were put on the territory of a male that had previously failed to mate (J. Høglund *et al.*). The result was an increase in the number of females that entered the territory of the less attractive male. However, the male still failed to mate with any female; so copying may attract females to inspect a male but clearly other criteria affect whether she will accept him.

There are alternative explanations for these observations. One possibility is that male attractiveness is not constant but varies from day to day, but this was ruled out in the study of sage grouse — there was an increase neither in the acoustic cues nor in the display rate on days that males copulated compared with days when they did not. Another possibility is that males increase their display rates once they have mated or if females are nearby. Females may merely be responding to the amplified display. Probably both copying and changes in male behaviour are important.

What benefits do females get from copying? If there is a cost involved in sampling and accurately assessing males, then copying may be a short cut to identifying a mate of high quality. Choice costs have not been measured in any lekking species but they undoubtedly exist.

1. Gibson, R.M. & Bradbury, J.W. *Behav. Ecol. Sociobiol.* **18**, 177–123 (1985).
2. Andersson, S. *Behav. Ecol. Sociobiol.* **25**, 403–410 (1989).
3. Wade, M.J. & Pruett-Jones, S.G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5749–5753 (1990).
4. Feller, W. *An Introduction to Probability Theory and its Applications* Vol. 1, 3rd edn (Wiley, New York, 1968).
5. Høglund, J., Alatalo, R.V. & Lundberg, A. *Behaviour* (in the press).
6. Clutton-Brock, T.H., Hiraiwa-Hasegawa, M. & Robertson, A. *Nature* **340**, 463–465 (1989).
7. Gibson, R.M. & Bradbury, J.W. in *Ecological Aspects of Social Evolution* (eds Rubenstein, R.I. & Wrangham, R.W.) 379–398 (Princeton Univ. Press, 1986).
8. Hamilton, W.D. *Nature* **228**, 1218–1220 (1970).

*Third International Conference of Behavioural Ecology, Uppsala, Sweden, 22–26 August 1990.

Female grouse spend an average of 2–3 mornings (and maybe evenings as well) at the lek observing males², time that could instead be spent feeding or starting to incubate eggs.

If copying serves to reduce the costs of mate choice, we should expect that the behaviour would be most obvious in those females less able to bear them. On these grounds, first-year females are good candidates for copying — younger hens are less experienced and tend to arrive later in the season. But there is no measurable increase in copying through the season on sage grouse leks (R. Gibson *et al.*).

A final twist to the story is the reverse behaviour observed in peacocks (M. Pet-

rie, Open University). Dominant females stayed on the lek after copulating and courted their mate again whenever another female approached. In all cases, 'post-coital' courtship stopped the secondary female's advance, forcing her to mate with less attractive males that had previously been avoided. Prevention of copying was observed only for the most attractive males. This looks like spiteful behaviour⁸; dominant females monopolize the best males to reduce the quality of other females' choice of mate. □

Andrew Pomiankowski is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

CHEMISTRY

Photochromic memory devices

Geoffrey J. Ashwell

MATERIALS which undergo a reversible change in their optical transmission characteristics when irradiated (the photochromics) offer potentially exciting applications as optical data-storage media; they may be switched at one wavelength and interrogated at another. But, as yet, the majority have the disadvantage of thermally reverting to the unswitched form. On page 658 of this issue¹, Liu, Hashimoto and Fujishima disclose a stable recording medium based upon the photoelectrochemical reduction of an azobenzene derivative. The photochromic properties of the azobenzenes are well established^{2,3}; *trans*-to-*cis* isomerization occurs when they are irradiated with an ultraviolet source with reversal occurring either thermally or at visible wavelengths. Liu *et al.*¹ now show that the less stable *cis* form (but not the *trans* form) may be electrochemically reduced to hydrazobenzene which is stable in an inert atmosphere (scheme 1), revealing the potential of the azobenzenes for use in optical memory devices.

There are numerous other classes of photochromic compounds, but the molecular changes that occur during irradiation divide the majority into five categories; photo-induced *trans*-to-*cis* isomerization (as with the azobenzenes¹⁻³ and thioindigos⁴); homolytic cleavage (for example, hexaphenylbisimidazole⁵); hydrogen transfer (such as in salicylidene anils⁶); valence tautomerism (for example, the fulgides⁷ and spiropyrans⁸); and charge transfer (such as the viologen salts⁹, Cu⁺TCNQ⁻ (ref. 9) and a class of zwitterionic materials of general formula D⁺—CH=C(CN)—C₆H₄—C(CN)⁻, where D⁺ is a heterocyclic cation¹⁰). In this final category, a requirement for optical bistability is that changes in the electronic distribution must be accompanied by modifications to either the molecular structure or the film structure.

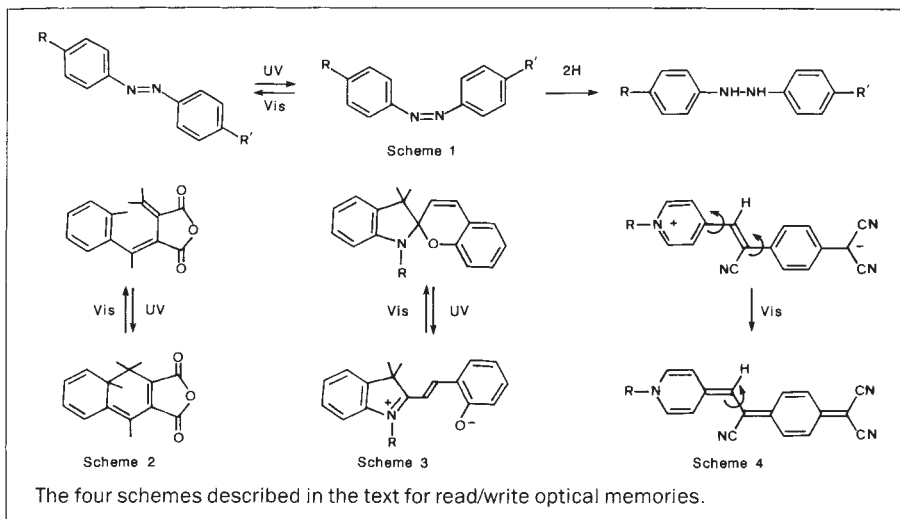
Photochromic materials have been studied for many years but so far their properties have not been exploited. Nonetheless, in the short term, the fulgides (scheme 2) and spiropyrans (scheme 3)

should find applications in security printing; they switch from being colourless to being coloured when irradiated with an ultraviolet source, the process being reversible at visible wavelengths, and some are fatigue resistant.

The spiropyrans⁷ also offer potentially exciting applications as components of a multifrequency memory device. Such devices require layers of different photochromic materials with non-overlapping absorption bands; by addressing each at a material-specific wavelength the layers may be switched independently of each other and thus provide the basis for three-dimensional data storage. But most photochromic materials have broad spectra when in the solid state whereas storing several bits per pixel requires materials with sharp absorption bands. Some π -bridged donor-acceptor systems satisfy this criterion and show promise in this area; Langmuir-Blodgett films of Z- β -(1-hexadecyl-4-pyridinium)- α -cyano-4-styryldicyanomethanide (scheme 4) and its quinolinium congener have sharp photochromic absorption maxima at 495 and 565 nm with half widths at half height of 22 and 27 nm respectively¹⁰. These bleach when irradiated at wavelengths which overlap the bands, the process being reversible in solution but irreversible in the films. Thus, they have a potential use as components of a multifrequency write-once/read-many (WORM) memory.

Although azobenzene is unsuitable as a multifrequency switching component, its absorption bands being too broad, Liu *et al.*¹ now demonstrate that it has an advantage over many other systems in so far as the read state (the hydrazobenzene) is both stable and erasable; it may be electrochemically oxidized to *trans*-azobenzene (the write state). They also show that the state of the film remains reversible over several hundred write/read cycles. Thus, this preliminary work is an important step forward in the development of an erasable photochromic memory and it is anticipated that such a device would have a storage density of 10⁸ bits per square centimetre. □

Geoffrey J. Ashwell is in the Centre for Molecular Electronics, Cranfield Institute of Technology, Cranfield MK43 0AL, UK.



1. Liu, Z.F., Hashimoto, K. & Fujishima, A. *Nature* **347**, 658–660 (1990).
2. Kano, K. *et al. Chem. Lett.* 421–424 (1980).
3. Yabe, A. *et al. Thin Solid Films* **160**, 33–41 (1988).
4. Mooney, W.F. *et al. J. Am. Chem. Soc.* **106**, 5659–5667 (1984).
5. Heller, H. *IEEE Proc.* **130**, 209–211 (1983).
6. Kawamura, S., Tsutsui, T., Saito, S., Murao, Y. & Kina, K. *J. Am. Chem. Soc.* **110**, 509–511 (1988).
7. Ando, E., Miyazaki, J., Morimoto, K., Nakahara, H. & Fukuda, K. *Thin Solid Films* **133**, 21–28 (1985).
8. Nagamura, T., Sakai, K. & Ogawa, T. *J. Chem. Soc., Chem. Commun.* 1035–1037 (1988).
9. Potember, R.S., Hoffman, R.C. & Poehler, T.O. *Johns Hopkins APL Technical Digest* **7**, 129–141 (1986).
10. Ashwell, G.J., Dawney, E.J.C., Kuczyński, A.P. & Szablewski, M. *Mat. Res. Soc. Symp. Proc.* **173**, 507–512 (1990).

The worm and the virus

F. E. G. Cox

INTEREST in the possible interactions between parasites and AIDS, particularly in Africa, has taken a new turn with a report in the latest issue of *Journal of Experimental Medicine*¹. In it, Jamal Khalife and colleagues describe the unexpected discovery that one of the surface antigens of the helminth worm *Schistosoma mansoni* contains an epitope that is identical to one of the regulatory proteins of the human immunodeficiency virus HIV-1. The finding bears on the controversy as to whether there is a causal relationship between the occurrence of schistosomiasis and AIDS, and also opens up a number of possibilities concerning the interactions between these pathogens.

Schistosomes evade the immune response by coating themselves with host molecules within a few days of penetrating the skin. Each adult worm carries a forest of acquired substances — including various blood groups, major-histocompatibility-complex molecules, immunoglobulins and albumin — which mask the worm antigens that might act as targets for immune attack². It is therefore difficult to determine exactly which antigens are specific to the parasite and which have been acquired from the host. The schistosome antigen described by Khalife *et al.*¹ is apparently worm-derived, and is a glycoprotein of relative molecular mass 170,000 (*M_r*, 170K) which appears on the surface of the invading larval schistosome after about three days, while it is still in the skin, and persists into the adult stage. It is of interest not only because it shares an epitope with one of the HIV-1 regulatory proteins, but because it also induces protection against challenge with *S. mansoni* in rats and induces a specific antibody response in humans.

The antigens of HIV-1 are coded for by nine genes, the best known being *gag* and *env*, which code for the structural elements (the core and envelope proteins), and *pol* which codes for enzymes. But there are also six regulatory genes, among them *vif* which codes for the virion infectivity factor, Vif, concerned with entry of free virus into the host cell³. Vif is a small protein which occurs on free virus particles, in the cytoplasm of infected cells and in the fluid surrounding them. But only a proportion of individuals infected by HIV-1 develop antibodies to Vif and its importance in AIDS is not clear⁴.

Khalife *et al.* have identified a 14-amino-acid epitope common to the HIV-1 Vif antigen and to the 170K antigen of *S. mansoni*. This peptide induces monoclonal antibodies that recognize both HIV-1 and the surface of *S. mansoni*, and it is also recognized by some, but not all,

individuals infected with either HIV-1 or *S. mansoni*. In rats, passive administration of the monoclonal antibody reduced the parasite load by 34–49 per cent, which is comparable with the best results obtained using antibodies against more conventional antigens⁵.

The immediate significance of this finding is not clear but its implications are

and tumour necrosis factor stimulate HIV-1 enhancers and thus act as cofactors in AIDS⁶. Similarly, cells infected with HIV-1 produce cytokines that are capable of modifying immune responses⁹, and there is every reason to suspect that the cytokines would affect *S. mansoni*. The existence of a common epitope also leaves open the possibility that the virus shares antigens with other parasites, such as the malaria organism, with which it appears to have a seroepidemiological association¹⁰.

The biological function and possible origins of the common Vif–170K antigen

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Schistosoma mansoni, one of three species of fluke which cause schistosomiasis.

wide ranging. *Schistosoma mansoni* and HIV-1 frequently occur together in Africa and there is some seroepidemiological evidence that there is a positive correlation between the two infections⁶. This has been disputed⁷, but the lack of correlation relates to the better known *S. mansoni* antigens and the structural antigens of the virus, and both viewpoints will have to be revised in the light of these new discoveries. Practical consequences include the possibilities that *S. mansoni* infections might actually inhibit the invasion of cells by HIV-1, or that HIV-1 infections might contribute to immune responses against *S. mansoni*. In this context, concepts of a cocktail vaccine against *S. mansoni* might have to be revised if it is found that one of the components actually interferes with HIV-1 — even though this might appear at first glance to be beneficial — because the balance between protection and pathology in schistosomiasis is so delicate.

The actual interactions between HIV-1 and *S. mansoni* are likely to be much more complex than they would first appear. *Schistosoma mansoni* infections induce the production of a number of cytokines including interleukin-1, interferon- γ and tumour necrosis factor, but interleukin-1

are also of interest. There are many similarities between HIV-1 and the simian immunodeficiency virus, including the presence in each of a Vif molecule¹¹. It would be very interesting to know more about a molecule that was common to the two organisms that are now pathogens before humans came on the evolutionary scene. Is it a schistosome structural or functional antigen, or was it acquired from its simian host? If so, could it possibly have come from a virus with which the host had come to terms? □

F. E. G. Cox is in the Division of Biomolecular Sciences, King's College London, Campden Hill Road, London W8 7AH, UK.

1. Khalife, J. *et al.* *J. exp. Med.* **172**, 1001–1004 (1990).
2. McLaren, D. J. *Parasitology* **88**, 597–611 (1984).
3. Haseltine, W. A. & Wong-Staal, F. *Sci. Am.* **259**, 34–42 (1988).
4. Weber, J. *Curr. Op. Immun.* **2**, 420–423 (1990).
5. Mitchell, G. F., Rogers, M. V. & Tiu, W. U. in *Vaccination Strategies of Tropical Diseases* (ed. Liew, F. Y.) 149–164 (CRC Press, Boca Raton, 1990).
6. Biggar, R. J. *et al.* *Int. J. Cancer* **35**, 763–767 (1985).
7. De Lima e Costa, M. F. F. *et al.* *Trans. R. Soc. trop. Med. Hyg.* **87**, 262 (1988).
8. Osborn, K., Kunkel, S. & Nabel, G. J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2336–2340 (1989).
9. Nakajima, K. *et al.* *J. Immun.* **142**, 531–536 (1989).
10. Biggar, R. J. *et al.* *Lancet* **ii**, 520–523 (1985).
11. Fultz, P. N. & Anderson, D. C. *Curr. Op. Immun.* **2**, 403–408 (1990).

OCEAN DRILLING PROGRAM

Getting the bit below the ridge

*Leg 132 engineering and scientific parties**

A LONG-STANDING objective of deep-ocean drilling has been to recover cores from the axial regions of active spreading ridges. These regions have been studied with sophisticated seismic and sonar instruments, camera sleds and manned submersibles. But until our voyage, the fresh, highly fractured character of the basalts from these regions has prevented the recovery of substantial amounts of

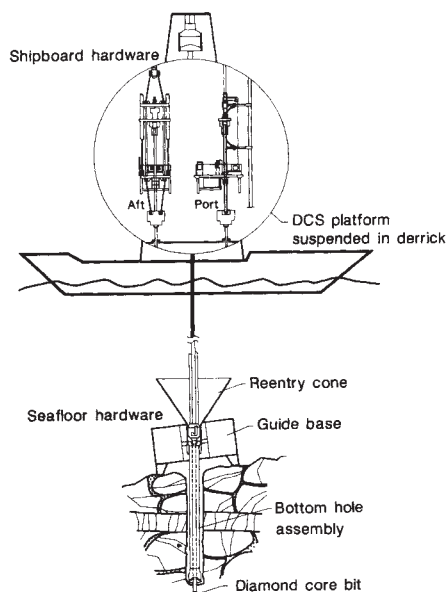


FIG. 1 Diamond coring system (DCS).

cored materials and only shallow holes have been penetrated. Now, using a specially adapted mining-type high-speed diamond coring system, we have drilled lava flows at the Sumisu Rift, an active submarine volcanic field near the Bonin island arc south of Japan.

Before this leg, ocean ridge formations were cored with conventional roller-cone bits about 10 inches in diameter, identical to bits often used in the oil industry. Recently erupted volcanic rock, however, is brittle and fractured. When rotary core bits were used in these unstable formations, the walls of the drill holes tended to collapse, binding the drill string and destroying the bit in short order. On land, such formations are drilled successfully with smaller-diameter diamond bits rotat-

ing at high speed with a low-level, precise drill-string weight on the bit. This action can be likened to that of a high-speed dental tool delicately cutting through the enamel of a tooth.

Such a system has now been developed for the Ocean Drilling Program for use aboard *JOIDES Resolution* (Fig. 1). Under the ship's 62-m derrick, a special mast 14 m tall supports an electric top drive used to rotate a tubing string that extends from the ship to the sea floor inside the ship's standard string of drill pipe. The diamond bit can be rotated at speeds of up to 540 revolutions per minute, 5–10 times faster than in conventional coring. Computer-controlled mechanisms keep the weight of the drill string at a uniformly low level. The paramount design objective for the diamond coring system was to adapt relatively fragile mining-drilling hardware to a configuration that could be used on a drill ship in deep water. The system needed to be strong enough to rotate a drill string reaching to the sea floor as much as 4.5 km beneath the ship, and delicate enough to core the fragile formations.

We tested the diamond coring system at a site (Fig. 2) located in 1,800 m of water on a small ridge of *en echelon* volcanic peaks in the central part of Sumisu Rift, a narrow, fault-bounded, extensional basin that has opened up during the past two million years or so to the west of the Izu–Bonin arc, south of Japan. The area has previously been surveyed extensively and sampled by submersible vessels. The ridge consists primarily of basalts with

LOCOMOTION

The price of being a snake

Roger C. Woledge and Nancy A. Curtin

ALTHOUGH most land animals use legs for moving about, it has been argued that this is inefficient compared with limbless locomotion, such as the lateral undulation of snakes, which avoids the costs at each stride of accelerating the legs and of moving the centre of gravity of the body vertically. This idea was supported by preliminary observations¹ on a garter snake (*Thamnophis sirtalis*), which suggested that the snake's oxygen consumption was low during movement. Now, however, it has been disproved by the more detailed experiments of Walton, Jayne and Bennett, published in *Science*², showing that the energetic cost of locomotion of snakes is similar to that for lizards.

Walton and colleagues used black racer snakes (*Coluber constrictor*) weighing

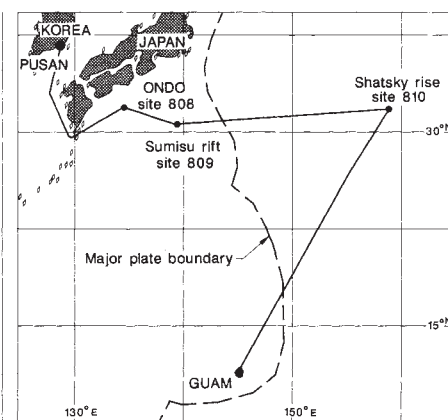


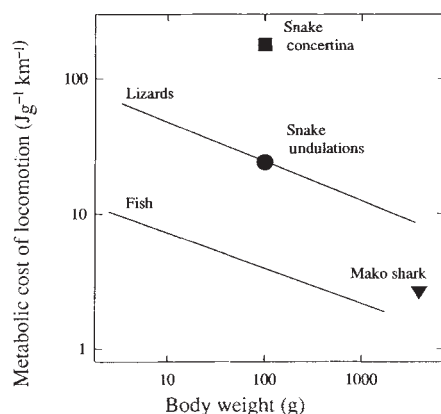
FIG. 2 Sites visited by Leg 132.

back-arc-basin geochemical characteristics, although some more-silica-rich and hybrid lavas have been found also. There are inactive hydrothermal vents at the northern end of the cluster of volcanoes.

Coring proceeded easily through fractured basalt and unconsolidated vitroclastic breccia. In the ten days of operation we drilled a hole 79 m deep with more than 60 per cent of the basalts cored being recovered. In contrast, for ODP Legs 106 and 109 on the Mid-Atlantic Ridge — two previous attempts to drill young, rift-axis basalt — hole stability was a continual problem and less than a tenth of the core was recovered. The hole was abandoned after only 50.5 m of penetration.

Future plans for the new system, which is currently being improved, include drilling the crest of the East Pacific Rise in the eastern Pacific and high-temperature hot springs off the coast of British Columbia. The diamond coring system will eventually be used to core certain types of sedimentary formations, as well as holes 2,000–3,000 m deep along thickly sedimented continental margins. □

* Michael A. Storms (Ocean Drilling Program, Texas A&M Univ.), James H. Natland (Scripps Inst. of Oceanography), Garrent W. Brass (Univ. of Miami), Glenn R. Brown (Univ. of Toronto), Jean-Baptiste Fay (Institut Français du Pétrole), Glen Foss (ODP, Texas A&M Univ.), G. Leon Holloway (ODP, Texas A&M Univ.), Steven P. Howard (ODP, Texas A&M Univ.), Dietmar Krehl (Eastman Christensen GmbH, West Germany), Ralf B. Luy (Institut für Tiefbohrtechnik, West Germany), Charles N. McKinnon, Jr. (Downey, California), Brian D. Mordaunt (DRECO Inc., Utah), Isabella Premoli-Silva (Univ. Milano), Frank Rack (ODP, Texas A&M Univ.), Daniel H. Reudelhuber (ODP, Texas A&M Univ.), William V. Sliter (U.S. Geological Survey), Robert J. Van Waasbergen (Scripps Inst. of Oceanography) and Masatake Zaitzu (Nippon Marine Enterprises Ltd, Japan).



The metabolic cost of locomotion of different animals. The cost of lizards⁷, which is typical of land animals³, is about eight times greater than for fish⁶. The snake, when using its more economical mode of locomotion, has a cost similar to that of other land animals, but much more than fish using lateral undulations of the body for locomotion.

probably because part of the metabolic cost has to be met from non-oxidative sources. At speeds up to 0.5 km h⁻¹, the rate of oxygen consumption increased linearly with speed. The slope of this relation gives the energetic cost of moving 1 km, and it turns out to be the same as that for lizards of similar body size. Concertina locomotion was almost ten times more costly than lateral undulation and maximum sustained speed was less. Thus in these snakes, at least, neither mode of locomotion is particularly economical.

Fish also move by lateral undulation and its metabolic cost is known for a wide range of fish species and sizes³. A recent example is the study of Graham *et al.*⁴, who used, among other sharks, a 3.9 kg Mako shark (*Isurus oxyrinchus*) swimming in a water tunnel immersed in the ocean. This fish swam for 36 hours in the tunnel at speeds of 0.5–1.4 km h⁻¹ (0.2–0.5 fish lengths per second) while oxygen consumption was monitored. Over this range of swimming speeds the rate of oxygen consumption is a linear function of speed. The slope of this relation gives the energetic cost of moving 1 km and, as shown in the figure, it is similar to the values for other fish. The striking feature is that the cost of locomotion in fish is considerably less than that for land animals, including snakes using lateral undulation.

In both of the studies (black racer snake and Mako shark), the mode of locomotion

was directly observed and gave values for the distance moved forward by the animal in one cycle of lateral undulation. In each case the value is about 0.3 m. Thus the much smaller energy cost for the fish moving 1 km is not because it uses fewer cycles of undulation than the snake does.

Why might fish locomotion be more economical than locomotion on land? The difference could be in the efficiency of energy transduction in the muscles themselves. But in the shark *Scyliorhinus canicula* the muscle efficiency is less⁵ than that of, for example, frog⁶, so the reason for the lower cost of locomotion in fish has to be found elsewhere. It could be that, in fish, less of the energy produced by the

muscles need be devoted to overcoming friction and supporting body weight than is the case for movements on land. Therefore, fish may use a smaller proportion of their muscle mass than land animals for locomotion at sustainable speeds. If this were the case, expressing the cost in terms of total body weight, rather than weight of active muscle, would make the fish seem to be more economical than they really are. □

Roger C. Woledge is in the Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK. Nancy A. Curtin is in the Department of Physiology, Charing Cross and Westminster Medical School, London W6 8RF, UK.

DEEP EARTHQUAKES

Metastable phases confirmed

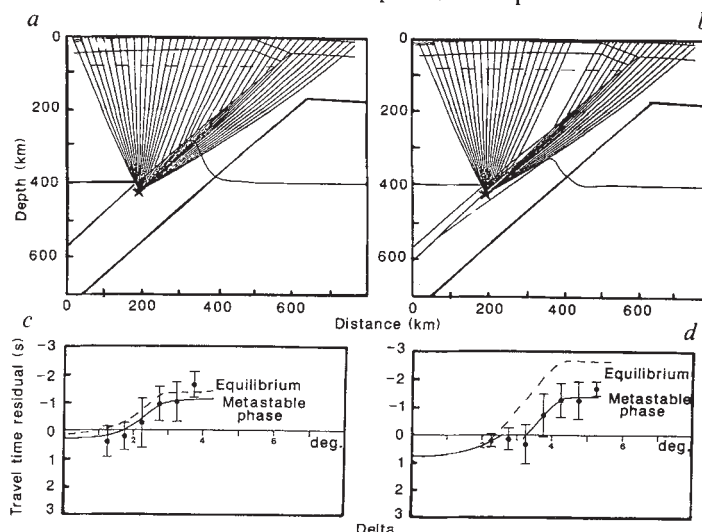
Robert J. Geller

THE mechanism responsible for deep focus earthquakes, which occur at depths of 300–700 km in subducting slabs, has been a long-standing problem in Earth sciences. But seismic results presented at a recent meeting* showed that a metastable phase of olivine exists in downgoing slabs. This metastable phase plays a key role in the mechanism of deep earthquakes.

Unlike shallow earthquakes, deep earthquakes cannot be attributed to frictional instabilities across a fault plane,

events with different geometries. In short, both deep and shallow earthquakes generate seismic waves with the same pattern of azimuthal variation.

An attractive possible faulting mechanism that generates the necessary seismic radiation pattern is the transformation-faulting hypothesis¹ (based on an analogy with faulting experiments at much lower pressures in ice and tremolite), according to which deep earthquakes result from the presence of a metastable phase under



Two-dimensional ray tracing for a model in which the 400-km discontinuity in the slab is governed by equilibrium kinetics² (a), or where a metastable phase exists³ (b). Both models fit the observed residuals for sources 300–350 km deep (c), but the data for sources 450–500 km deep (d) support the existence of a metastable phase. (Courtesy T. Iidaka and D. Suetsugu.)

because of the high frictional resistance to sliding at depth. On the other hand, a careful analysis of seismic waves generated by deep earthquakes (H. Kawakatsu, Geological Survey of Japan and K. Kuge, Tokyo University) confirms that a sudden impulsive phase change can be ruled out as the primary physical mechanism. The same analysis also showed that a deep earthquake which appeared to have an atypical fracture mechanism was simply the superposition of several faulting

stress in the subducting slab. Ambient shear stresses in the slab trigger and localize an overdriven phase change in a growing fault zone; the ambient shear stresses are then released by faulting along this zone, which is weakened by reduced grain size and latent heat.

With increasing depth, the stable phase of olivine, the most abundant mineral in the upper mantle becomes first β -spinel and then γ -spinel. The first of these phase transformations is generally thought to be responsible for the seismic discontinuity found at a depth of 400 km in the

- Chodrow, R.E. & Taylor, C.R. *Fedn. Proc.* **32**, 422 (1973).
- Walton, M., Jayne, B.C. & Bennett, A.F. *Science* **249**, 524–527 (1990).
- Schmidt-Nielsen, K. *Scaling: Why is Animal Size So Important?* 187 (Cambridge Univ. Press, 1984).
- Graham, J.B., Dewar, H., Lai, N.C., Lowell, W.R. & Arce, S.M. *J. exp. Biol.* **151**, 175–192 (1990).
- Curtin, N.A. & Woledge, R.C. *J. Physiol.* **407**, 74p (1988).
- Kushmerick, M.J. & Davies, R.E. *Proc. R. Soc. B* **174**, 315–353 (1969).
- John-Alder, H.B., Garland, T. & Bennett, A.F. *Physiol. Zool.* **59**, 523 (1986).
- Beamish, F.W.H. in *Fish Physiology* Vol. 7 (eds Hoar, W.S. & Randall, D.J.) 101–187 (Academic, New York, 1978).

* Western Pacific Geophysics Meeting at Kanazawa, Japan, 21–25 August 1990.

mantle. But in subducting slabs the cooler temperatures make the β -spinel phase barely favourable to olivine²; the over-driven exothermic transformation of olivine to the γ -spinel phase is the likely candidate for the transformation faulting model of deep earthquakes (S. Kirby, US Geological Survey).

Laboratory experiments³ on the $\alpha \rightarrow \gamma$ transformation in Mg_2GeO_4 , which is a low-pressure analogue of olivine, show that, under appropriate conditions of temperature and non-hydrostatic stress, the transformation from the metastable α phase to the γ phase can induce faulting. Moreover, these experiments exhibit 'anticrack' microstructures that provide insights into the underlying mechanism of deep earthquakes^{4,5}. Similar experimental results on the $\alpha \rightarrow \beta$ transformation have now been found at a much higher pressure (14 gigapascal) in natural olivine (H. Green and T. Young, University of California, Davis; D. Walker and C. Scholz, Columbia University), greatly strengthening the case for this model.

Independent evidence for the existence of the hypothesized metastable phase within subducting slabs is needed to confirm the mechanism. This should be forthcoming from observations of the displacement of the 400-km discontinuity. Within the subducting slab, this discontinuity should be raised by about 100 km if the material is at equilibrium⁶, but if there is a metastable phase the discontinuity would be deflected sharply downwards⁷.

Although seismological observations offer the best prospects for confirming or refuting the existence of a metastable phase, high-quality data and a carefully chosen source-receiver geometry are required. Arrival-time data from about 60 stations in the Tokai and Kanto districts for about 30 large earthquakes in the subducting Pacific plate at depths from 300 to 500 km (T. Iidaka and D. Suetsugu, Tokyo University) show that the travel-time residuals are consistent with the existence of a metastable phase, and inconsistent with a model in which the phase transition obeyed equilibrium thermodynamics (see figure). Although the details are somewhat model dependent, the case for the existence of the metastable phase in the downgoing slab appears strong; the case will be further strengthened if amplitude and waveform studies confirm this result.

A direct seismological observation of the 670-km discontinuity inside the downgoing slab (C. Wicks and M. Richards, University of California, Berkeley), using short-period data from earthquakes in the Tonga and Izu-Bonin arcs and an array in Australia, show that it is not appreciably deflected downward, which suggests that this discontinuity is a phase change rather than a chemical discontinuity. This result

has important implications for the geodynamics of the mantle and its geochemistry. It is also consistent with the idea that deep earthquakes cease below 670 km, because the metastable phase ceases to exist¹.

The fact that several independent lines of evidence — high-pressure faulting experiments, seismological observations of the fine structure of slabs, and seismological studies of deep earthquake source mechanisms — fit together so well is a highly encouraging sign that Earth scientists are on the right track. Nevertheless, controversial points remain (Kirby and Green, personal communications). Green and colleagues report that the stable phase forms as anticracks everywhere, not just in a long, thin zone, and argue that the faulting is a plastic instability accompanying an exothermic reaction that is otherwise analogous to brittle failure, in that faulting does not develop before the onset of the instability.

On the other hand, on the basis of his

EVOLUTION

Between biology and culture

Rory Howlett

THE success of molecular biology lies, in part, on the ability it gives us to hang observations made at the molecular level on the conceptual framework provided by darwinian evolution. So it was appropriate that as one of the events in the centenary year of Cold Spring Harbor Laboratory, a centre of excellence in molecular biology, a meeting* should be held to address one of the outstanding problems in evolution — to what extent there can be a useful interplay of ideas between biologists and those concerned with language and culture.

At first sight, molecular and evolutionary biology on the one hand, and linguistics and archaeology on the other, are distinct disciplines that have developed along quite different paths. But there are some questions, such as the origin of similarities, which they share in common, and there has long been the hope that analogies between the modes of genetic and cultural change (for example, the analogy between units of genetic inheritance, genes, and units of cultural inheritance, memes¹) will allow a common approach.

Similarity, whether it be between DNA sequences, behaviours, languages or cultural entities, may come about through a number of different mechanisms. The three most commonly encountered in evolution are descent from a common ancestor; convergence, by which traits with independent origins converge under the influence of similar selection pressures to fulfil similar functions; and borrowing,

recent work in ice, Kirby argues that the transformation flaws interact before failure by transformation faulting; he thus views transformation faulting in ice, germanates and metastable olivine as the result of strongly exothermic metastable phase changes governed by the same underlying physics. This controversy indicates the need for a more quantitative understanding of the physics of deep earthquakes. □

Robert J. Geller is in the Department of Geophysics, Faculty of Science, Tokyo University, Bunkyo-ku, Tokyo 113, Japan.

1. Kirby, S. H. *J. geophys. Res.* **92**, 13789–13800 (1987).
2. Akaogi, M., Ito, E. & Navrotsky, A. *J. geophys. Res.* **94**, 15671–15685 (1989).
3. Burnley, P. C. & Green, H. W. *Nature* **338**, 753–756 (1989).
4. Green, H. W. & Burnley, P. C. *Nature* **341**, 733–737 (1989).
5. Fröhlich, C. *Nature* **341**, 687–688 (1989).
6. Schubert, G., Yuen, D. A. & Turcotte, D. L. *Geophys. J. R. astr. Soc.* **52**, 705–735 (1975).
7. Sung, C.-M. & Burns, R. G. *Earth planet. Sci. Lett.* **32**, 165–170 (1976).

by which traits are directly transferred from one line to another (horizontal DNA transfer would be an example at the molecular level). A main aim of the meeting was to assess the relative importance of these mechanisms in different evolutionary fields.

In one specific instance², it has been argued that convergent evolution to function at acidic pH accounts for the similarities between the stomach lysozymes of ruminants and those of colobine monkeys, such as the langur, which have fermentative foreguts. But in molecular evolution it is usually assumed that sequence convergence is rare (R. Doolittle, University of California, San Diego). This follows from the observation that, of the vast number of possible protein sequences, only a vanishingly small proportion exist. Doolittle therefore proposes that similar sequences are most likely to have come about either by chance or by common ancestry. Nevertheless, there is a whole growth industry based on the need to develop computer algorithms for the detection of sequence similarities, and to formulate statistical methods for determining whether similarities are such that they are unlikely to have come about by chance.

To underplay further the importance of convergence (and borrowing) at the molecular level, there are the points that the number of new proteins which have arisen during the course of evolution is unlikely to be more than 2,000–4,000, and that most new gene functions come about because of the release from functional constraints of temporarily redundant

* *Evolution: Molecules to Culture*, Cold Spring Harbor Laboratory, 24–27 September 1990.

copies of genes following, for example, duplication events (C. Chothia, Medical Research Council, Cambridge). Similarly, proteins can alter their function through step-by-step changes so as to maintain structural stability. A good example is the evolution of the crystallins, structural proteins found in eye lenses, from enzymes (see an earlier discussion in *News and Views*¹). An alternative pathway by which new proteins can evolve is by the stitching together of 'modules', leading to proteins with novel functions.

Although natural selection has for the most part been responsible for the evolution of protein function, most changes at the DNA level are now accepted to be neutral in that they are neither detrimental nor advantageous to the organisms that bear them. Using population genetics theory, neutralists (such as M. Kimura, National Institute of Genetics, Japan) argue that the rate of molecular evolution is limited by the mutation rate. By contrast, neodarwinists contend that the evolution of primary interest is that which gives rise to adaptation, and that the arising of new mutations can be seen as the blank block of stone which is reduced by selection to create adaptation (R. Dawkins, University of Oxford).

Recently there have been enormous methodological and statistical advances in the analysis of adaptation and biological diversity, whether it be at the level of morphology or behaviour. The comparative method in particular has been invaluable in identifying the evolutionary forces responsible for moulding adaptation (P. Harvey, University of Oxford). A key step forward has been the formulation of statistical methods for analysing correlation between different traits such as brain and body size, and ecological factors, while controlling for the effects of phylogeny. In linguistics, the comparison has been used to reconstruct the relationships between extant languages to build language family trees; but there is a need for formal models to tackle such questions as the rate of divergence between languages with common roots, and allow researchers to investigate the mechanisms by which such divergence takes place (M. Pagel, University of Oxford).

Borrowing

Because information can be transferred between individuals by learning, there is plenty of scope for borrowing in cultural evolution. This 'cultural transmission' of information is not unique to humans, and one striking example in animal behaviour comes from studies of the songs of birds, such as the European starling, introduced into New Zealand in the last century (P. Jenkins, University of Auckland). Introductions to isolated islands provide an unparalleled opportunity to unravel the effects of genetic and cultural pro-

cesses taking place during song ontogeny.

Birds produce song using a bipartite organ known as the syrinx, the two parts of which can generate sounds independently². Jenkins found that starlings are able to mimic the songs of two other birds simultaneously using their 'two voices', but that the actual phrases borrowed are apparently chosen to match an underlying species-specific song template which presumably has a genetic basis.

But how far can one push the analogy between genetic and cultural evolution to derive insights about the way languages and cultures evolve? For languages, the fact that, at the coarse level at least, boundaries between them often correspond with discontinuities in the genetic constitution of human populations (L. Cavalli-Sforza, Stanford University) implies that genes and languages to some extent share a common evolutionary history (see also the findings described in *News and Views* two months ago³).

Protolanguages

Optimists take the view that it may even be possible to reconstruct protolanguages ancestral to extant language families (S.A. Starostin, USSR Academy of Sciences) — in the case of Nostratic languages, for example, by the application of statistical methods to assess phonetic correspondences between language subgroups, based on objective linguistic genealogical trees. Nevertheless there are great difficulties to be overcome; some, such as loss of information because of the rapid change of divergent languages, may be insurmountable.

Indeed, whereas common ancestry may account for many 'global' aspects of human culture such as speech, C. Renfrew (University of Cambridge) argues that the types of methodologies generally applied by evolutionary biologists are unlikely to be able to cope with the realities of cultural evolution which include the widespread occurrence of independent invention and borrowing. Renfrew even questions the extent to which it is possible to recognize and define similarities in human culture, let alone quantify them.

So did the meeting result in a meeting of minds? Sad to say, my impression is that bringing together researchers from such disparate disciplines as molecular biology and linguistics is rather like trying to mix two immiscible liquids. When shaken together, the liquids may for a moment seem to dissolve into one. In time, however, they separate out again. □

Rory Howlett is an assistant editor of Nature.

Deep acupuncture

SMALL foreign objects can be quite mobile in the body. Embedded shrapnel splinters or porcupine quill tips can slowly travel through the tissues, ratcheted along by the ceaseless movements of the body acting on their barbed or asymmetric structure. They may ultimately emerge far away from the original site of entry.

On a microscopic scale, a small foreign body must be very blunt. It does not pierce or damage the cells it encounters, but merely noses them aside: this is why a hypodermic needle does so little damage. So Daedalus is now inventing telemicrosurgery. A little barbed metal needle is inserted under the patient's skin, and is guided to its target in the body by vibromassage pads applied at strategic sites. The needle is steered by altering the direction of vibration and the relative phases of the massage pads, while its progress is monitored by X-ray. With a sufficiently high vibrational frequency, the micro-missile could reach its goal in days or hours rather than the years that a porcupine quill tip may take on its odyssey.

Most medical applications will use a hollow needle. It may be filled either with a potent drug too dangerous to release into the body at large, or pure vacuum: sealed in either case by some low melting biocompatible wax. When it reaches its target, an alternating magnetic field will be applied over the site to warm the metal needle by induction. The wax will melt, either releasing the drug exactly at its desired site, or sucking a biopsy sample of a few cells back into the needle. It can then be steered back to the surface, for recharging or examination of its catch.

Thus the surgeon will be able to reach any site in the body without making an incision. With stronger induction-field heating, the micro-needle may even be useful in minor surgery — perhaps warming a region of tissue to promote healing, burning open a drainage path for an internal site of infection, or killing encysted parasites. In particular, it will transform neurology. The little needle could safely enter the most crucial nerve trunk, nosing its way among the mass of nerve fibres without doing any harm. It could identify each fibre it reaches by firing it: perhaps by mere contact, or maybe by the alternating current induced by a weak low-frequency magnetic induction field. When a particularly troublesome nerve is encountered (such as those which cause persistent neuralgia, or the phantom-limb pain of amputees) it could be specifically killed. Just turn the induction field up to maximum, and the suddenly heated needle will cauterize the offending nerve fibre, without affecting even its close neighbours. No conventional surgery could be so precise!

David Jones

1. Dawkins, R. *The Selfish Gene* (Oxford Univ. Press, 1976).
2. Stewart, C.B., Schilling, J.W. & Wilson, A.C. *Nature*, **330**, 401–404 (1987).
3. Doolittle, R.F. *Nature* **336**, 18 (1988).
4. Suthers, R.A. *Nature* **347**, 473–477 (1990).
5. Barton, N.H. & Jones, J.S. *Nature* **346**, 415–416 (1990).

Domestic mercury pollution

SIR—Mills in his recent Scientific Correspondence¹ identified a previously unsuspected source of atmospheric mercury and speculated on its toxicity. Here, I report another novel source of mercury pollution in predominantly domestic settings, where it has hitherto been thought a rarity². I believe this is the first mention of widespread domestic exposure to mercury vapour.

Mercury metal is used in Latin American and Caribbean communities for occult purposes. It is sold in shops called botanicas which stock medicinal plants, magical medicines, incense, candles and perfumes³. These and other religious articles are dispensed to adherents of syncretistic Afro-Caribbean/Latin American religions, such as Santería and Voodoo, as well as to the general public.

In a survey of 115 botanicas in 13 cities with large Hispanic populations in the United States and Puerto Rico, 99 were found to sell mercury. A brief survey of botanicas in Mexico, Colombia and the Dominican Republic revealed that they too dispense mercury.

Botanicas typically dispense mercury in gelatin capsules, or occasionally glass vials. Of 41 samples purchased, the median amount dispensed was 8.5 g, with a mean of 9.0 g, a range of 1.5 to 31.3 g and a modal cost of \$1.00. Of 28 botanicas visited in New York City, 13 prescribed that mercury be sprinkled on the floor or mixed with soap and water used to mop the floor, to rid the house of evil influences or for other purposes. Some botanicas suggested repeated application at intervals of three days to a week, until the desired result is attained. One shopkeeper recommended placing mercury in an open container with a magnet. Any of these procedures would liberate mercury vapour directly into the room's atmosphere.

Several shopkeepers recommended that mercury should be carried on the person, kept in containers in the house, placed in bath water, or mixed with perfume, soap solutions or ammonia and camphor. Others prescribed placing mercury in devotional candles.

Mercury also presents an occupational hazard. My colleague twice observed shopkeepers spill mercury without removing or chemically neutralizing it. One proprietress said that she intentionally scattered mercury about her botanica to "bring good things", as well as adding a bit to each prescription she dispensed.

It is evident that the concentration of mercury vapour in room air will vary widely, depending on such factors as quantity dispensed, amount used, frequency of application, type of floor surface, air temperature, volume of room, height above floor level and ventilation. It is equally

apparent that the proposed $1 \mu\text{g m}^{-3}$ "upper limit for long-term exposure to mercury vapour in the air" cited by Mills¹ could easily be exceeded by repeated applications in a small apartment.

The most likely victims of these practices are young children, exposed to the highest levels of vapour as they sleep, crawl and play on contaminated floors^{4,5}. Mercury vapour enters the fetus via the placenta, and the infant via breast milk⁶. One result of mercury vapour's neurotoxicity on penetrating the blood-brain barrier⁷ is the subtle personality change called erethism. It seems likely that children are the principal victims of mercurial erethism, which is characterized by hostility, withdrawal, tendency to resent being observed, quick temper, loss of self confidence and loss of memory^{8,9}. Sunderman's advice⁸ of "giving careful consideration to intoxication from undetected exposure to mercury... in patients

encountering depressions, behavioural and neurological disorders", is worth heeding, particularly in the evaluation of emotionally disturbed children.

There seems ample justification for a programme to measure mercury vapour levels and to test exposed individuals. Sociological research is also required to develop an effective health-education programme for botanica owners and their clients.

ARNOLD P. WENDROFF

Occupational Therapy Program,
Columbia University,
630 West 168th Street,
New York, New York 10032, USA

1. Mills, A. *Nature* **346**, 615 (1990).
2. McNeil, N. I., Oliver, R. E., Issler, H. C. & Wrong, O. M. *Lancet* **i**, 270 (1984).
3. Murphy, J. M. *Santería: An African Religion in America*. (Beacon, Boston, 1988).
4. Curtis, H. A., Ferguson, S. D., Kell, R. L. & Samuel, A. H. *Archs Dis. Childhood* **62**, 294 (1987).
5. Hardy, H. L. & Finkel, A. J. *Hamilton & Hardy's Industrial Toxicology* 4th edn 101 (Wright, Boston, 1983).
6. Battigelli, M. C. in *Environmental and Occupational Medicine*. (ed. Rom, W. N.) 453 (Little Brown, Boston, 1983).
7. Choi, B. H. *Progr. Neurobiol.* **32**, 449 (1989).
8. Sunderman, F. W. *Ann. Clin. Lab. Sci.* **18**, 99 (1988).

More ellipses

SIR—Klaczko and Bitner-Mathe in their Scientific Correspondence¹ reported that a large portion of the wing outlines of several species of *Drosophila* could be described as elliptical arcs and hence by only two parameters. But the fit can be improved by two methods, both using Fourier analysis, which would allow more information to be extracted than the length to width ratio of the ellipse.

In the first method², the radius function (the distance of the perimeter from an arbitrary central point as a function of the angle subtended from a chosen axis) is decomposed to a set of harmonics using Fourier analysis. The arc can then be described as precisely as is required by a set of $2n$ coefficients, where n is the number of harmonics. Unfortunately, this simple method gives different results depending on the selection of the central point and axis, so it is of limited value in comparing shapes with few homologous points of reference.

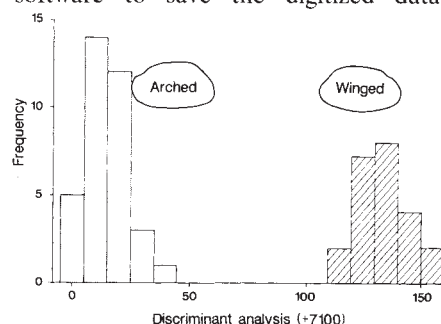
The second method³ uses two axes and the coefficients are independent of the size, position and orientation of the object. The arc is first smoothed by replacing it by short line segments pointing at either 0, 45, 90, ... or 315 degrees to a chosen axis. A Fourier series can then be found describing the X and Y projections of a point on this new arc to any required accuracy.

If one or more coefficients is set to zero then primitive shapes are obtained, of which the simplest are ellipses. Each harmonic in the Fourier series can be represented as a phasor, each of which has an elliptical locus, and the sum of the loci is an epicyclic path which approximates

the outline. The ellipses can be rotated and scaled into a standard position so that the coefficients depend only on the shape of the outline and not on the orientation in which the object is recorded.

Fourier coefficients are difficult to interpret geometrically because information about a single feature becomes confounded across several harmonics. Nevertheless, the distinctiveness of a feature on the outline will be associated with the eccentricity of certain elliptic loci. A regular partitioning of the outline into $n+1$ arcs should affect the size of the n th ellipse. Because there are two series there are $4n$ coefficients which can be analysed to estimate the similarity of the shape of different outlines such as wings.

Both methods are limited by the accuracy of digitization and number of points in the outline. Errors in the digitization are more likely to affect the higher harmonics. I recorded the valve outlines of freshwater mussels using a Graphtek digitizing pad model KD4030 connected to an IBM personal computer (this could be better done with video camera equipment). I have written Turbo Pascal v4.0 software to save the digitized data,



Discriminant analysis for the hyriid mussel *A. jacksoni*.

calculate the Fourier coefficients and display the reconstructed outlines.

The hyriid mussel *Alathyria jacksoni* has either a curved anterior-posterior ridge (arched) or a straight ridge (winged), depending upon environmental factors. Using the Fourier coefficients I calculated a discriminant function which distinguished perfectly between the two growth forms (see figure). I was surprised by this outcome as to distinguish the two forms by eye requires more information than just the valve outline; the posterior ridge and muscle scars must also be examined. The mussels could be separated into winged and arched groups by principal components analysis.

ROSS CORKREY

Department of Zoology,
University of Adelaide,
GPO Box 498, Adelaide,
South Australia 5001, Australia

1. Klaczko, L.B. & Bitner-Mathe, B.C. *Nature* **346**, 321 (1990).
2. Kaesler, R.L. & Waters, J.A. *Geol. Soc. Am. Bull.* **83**, 1169–1178, (1972).
3. Kuhl, F. & Giardina, C.R. *Comput. Graphics Image Process.* **18**, 236–258 (1982).

Hot springs and the origin of life

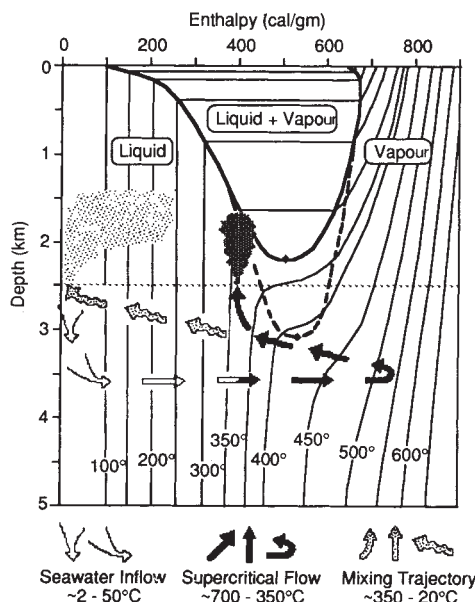
SIR—Miller and Bada¹ argue that prebiotic synthesis leading to the emergence of life could not have occurred in submarine hot springs. They state that “there is no thermal gradient until the waters at ~350 °C exit the vents and are rapidly quenched to deep ocean temperatures. Thus, the only thermal gradient in the system is in the area near the vent opening, which is where the proposed organic synthesis and further steps in the origin of life would have to take place.”

This description of the nature of submarine hot springs is incorrect. In addition to the high-temperature vents they describe, there are the ubiquitous low-temperature vents resulting from a distinct fluid trajectory (see figure). Hot-spring fluid samples collected on the Galapagos Rift in 1977 (ref. 2) revealed that the low-temperature vent fluids are produced by mixing within the crust of a high-temperature component, which last equilibrated with the rock at 350 °C, and ambient sea water which saturates the surrounding fractured crust. These Galapagos-type vents always accompany high-temperature smoker vents³.

Miller and Bada cite the suggestion of Turekian and Cochran⁴, based on radionuclides in clams from the vents, that the fluids remain within the crust at extreme temperatures for tens of years. Two findings refute this.

First, retaining large volumes of high-temperature fluids⁵ within the crust is problematic. Thermal budgets⁶ of submarine hot springs show that heat is

extracted from the magma by the advance of a cracking front⁷ into the magma body (see figure). Carbon in the magma and the seawater equilibrate isotopically⁸ at 640 to 720 °C in the chemical environment produced by the reaction of superheated water with the hot rock. Buoyant forces acting on these supercritical low-density/low-viscosity fluids inevitably lead to their rapid ascent, and to rapid mixing and quenching of hot fluid with cold sea



Fluid trajectories in submarine hot springs. The data are for fresh water with the critical point of sea water drawn in; 2,500 m is a typical sea-floor depth for mid-ocean ridge hot springs. A typical trajectory into the system, and two alternative pathways out, is shown on the enthalpy–depth diagram. The hot-spring model proposes that a fraction of the supercritical fluid produced at low temperature, by contrast to the fraction which rises rapidly without mixing to emerge as the “smoker” vents referred to by Miller and Bada. (From ref. 14).

water drawn in from the surrounding permeable crust.

Second, Kadko and Moore⁹ conclude from extensive radionuclide data on actual hot-spring fluid samples that the residence time of the fluids is much shorter. Their data are compatible with rapid heating and quenching.

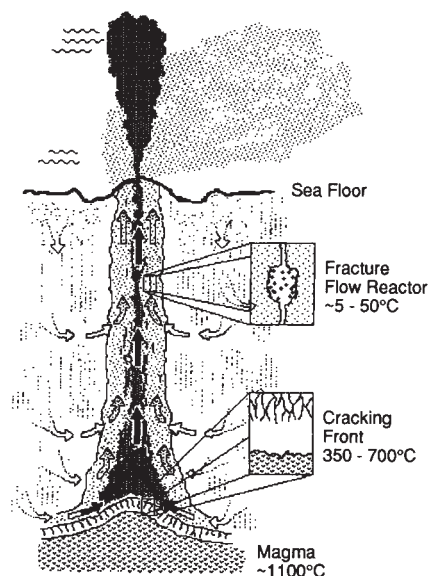
The presence of the mixing trajectory in the hot springs obviates the objections of Miller and Bada. Rapid quenching allows the conversion of thermal energy into high-energy bonds of the appropriate molecules which survive at low temperatures, lower than 50 °C for example.

Miller and Bada's claim that polymerization of amino acids into polypeptides will not occur in the presence of liquid water is contradicted by the work of Yanagawa and colleagues¹⁰, who report the production of polymers with peptide bonds in a modified sea water containing amino acids held at hot-spring temperatures for several weeks.

Miller and Bada's proposed analysis of the vent fluids searching for abiotically synthesized organic molecules would probably be futile in the relevant low temperature vents. These fluids contain up to 10^8 – 10^9 bacteria per cm^3 (ref. 2).

Concentrations of biologically produced organic compounds would mask any abiogenic component.

Joyce's News and Views comment¹¹ on Miller and Bada's paper, that “it is almost inconceivable” that the first living system was not a heterotroph, overlooks evidence from analysis of nucleotide sequences in 16S ribosomal RNA in modern organisms, which strongly suggests that a chemosynthetic thermophile



was the closest descendant of the earliest self-replicators¹². An enormous standing crop of these bacteria live within fracture “flow-reactors” in the warm upper reaches of these convective systems today. These natural analogues of Eigen's evolution reactor¹³ provide an optimum site for the emergence of these chemosynthetic thermophiles from non-living systems¹⁴.

JOHN B. CORLISS

Computer Systems Research Facility:
Code 935,
NASA Goddard Space Flight Center,
Greenbelt, Maryland 20771, USA

1. Miller, S. L. & Bada, J. L. *Nature* **334**, 609–611 (1988).
2. Corliss, J. B. *et al.* *Science* **203**, 1073 (1979).
3. Spiess, F. N. *et al.* *Science* **207**, 1421–1433 (1980).
4. Turekian, K. K. & Cochran, J. K. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6241–6244 (1986).
5. Macdonald, K. C. *Earth planet. Sci. Lett.* **48**, 1–7 (1980).
6. Cann, J. R. & Strens, M. R. *Nature* **298**, 147–149 (1982).
7. Lister, C. R. B. *Geophys. J. R. astr. Soc.* **39**, 465 (1974).
8. Welhan, J. A. & Craig, H. in *Hydrothermal Processes at Sea Floor Spreading Centers* (eds Rona, P. A. *et al.*) (Plenum, New York, 1983).
9. Kadko, D. & Moore, W. *Geochim. cosmochim. Acta* **52**, 659–668 (1988).
10. Yanagawa, H. *et al.* *Origins Life Evol. Biosph.* **18**, 179–207 (1988).
11. Joyce, G. *Nature* **334**, 564 (1988).
12. Woese, C. R. *Microbiol. Rev.* **51**, 221–271 (1987).
13. Eigen, M. *Naturwissenschaften* **58**, 465–523 (1971).
14. Corliss, J. B. *Life is a Strange Attractor* (Ariel, Santa Cruz, in the press).

Sirius and the colour enigma

SIR—Some ancient Babylonian, Graeco-Roman, and mediaeval texts¹ have suggested that Sirius might have been red in the past. In particular, Ptolemy (AD 150) classified Sirius as *hipokkiros* (reddish) with stars like Arcturus, Aldebaran, Pollux, Antares and Betelgeuse, all of which have B-V colours higher than 1. But the redness of Sirius is still contentious². We have found a new clue from Chinese records, and suggest a new possible explanation of the hypothetical redness of Sirius about 2,000 years ago — the passage of a small interstellar cloud across our line of sight to Sirius.

Several Chinese records referring to Sirius as a white star have been quoted to challenge the results of the colour change³. But during a recent visit to China, one of us found in the Han dynasty history book (Shiji) of the famous historian and astronomer Sima Qian (145–87 BC) a record, shown and translated in the figure. Following the practice of the time, the astronomical observation is followed by astrological predictions. The astronomical content of Shiji has been generally reliable in these cases (comets' or planets' motions) when direct confirmation can be performed. As for the rest of the compilation work of Sima Qian, no definitive date can be attached to the record, but it does show that a colour change in Sirius may have been noticed by Chinese astronomers more than 2,000 years ago.

It is unquestionable that Sirius was red when Sirius B passed through the red giant phase. But stellar evolution is not rapid enough that this transition could have taken place in human history.

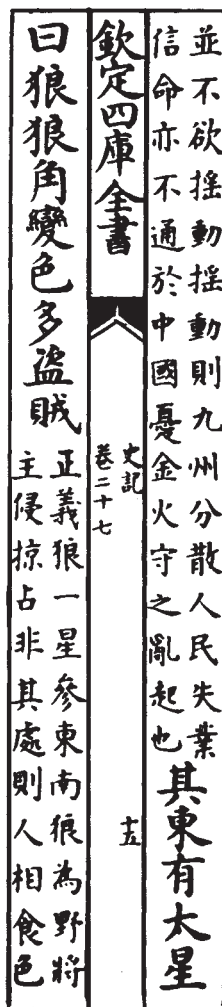
We think that the reddening of Sirius could have been caused by an interstellar cloud travelling between it and the Sun. The Bok globules, discovered by Bok in the 1940s, are the smallest dark nebulae that have been observed⁴. They are between 0.01 and 0.1 parsecs in diameter and between 0.1 and 1.0 solar masses. They are usually seen as dark regions in front of bright emission nebulae, but whether or not they are actually associated with the nebulae is uncertain. About 200 globules are known within 150 parsecs from the Sun, so the presence of a small globule in front of Sirius should have a non-negligible probability.

If we suppose that Sirius used to have a B-V colour of 1, putting it among the stars called *hipokkiros* by Ptolemy, the required reddening must have been $E(B-V) = 1$. Adopting the standard relation of Spitzer⁵, $N(H)/E(B-V) = 5.4 \times 10^{21}$, we find that the absorbing cloud must have a column density of $5.4 \times 10^{21} \text{ cm}^{-2}$. For a gas-to-dust ratio⁶ of $A/E(B-V) = 3.0$, we find that the extinction caused by the small

cloud is $A = 3 \text{ mag}$, giving Sirius an apparent magnitude of about +1.5. This is, of course, much fainter than it is now, but still a first magnitude star for the ancients.

If a globule of typical diameter of 0.02 parsecs lies just in front of Sirius, say at a distance of 2.5 parsecs from the Sun, its apparent dimension on the sky would be around 25 arcmin. Its mass then can be

Chinese record from the Han dynasty (second century BC) history book *Shiji*, chapter 27, *Book of Heavenly Bodies* compiled by Sima Qian, referring to a colour change in Sirius. Text from the Qing-ding-si-kuan-shu (ed. Zhong Hua Shu Ju, 1959/9) translates literally as: At East/ there is/ big/ star/ called/ Wolf/ Wolf/ horn/ changes / colour/ / many/ thieves/ robbers/. (Wolf is the Chinese name for Sirius.) The text corresponds to the large heavy characters in the first and third columns, to be read vertically and from right to left.



calculated as: $M = N (\text{cm}^{-2}) D^2 (\text{pc}^3) 4.2 \times 10^{21}$ solar masses, ~ 0.01 solar masses. Sirius has a proper motion of 15 km s^{-1} (or $1.3 \text{ arcsec yr}^{-1}$). If the globule were motionless, Sirius would have taken about 1,100 years to travel behind it. This time could be shorter depending on the proper motion of the cloud, but the timescale is however compatible with observational reports.

Note that the colour change would in this case have occurred progressively and would probably not have been noticeable by a human eye in a human lifetime,

- Schlosser, W. & Bergmann W. *Nature* **318**, 45 (1985).
- McCluskey, S.C. *Nature* **325**, 87 (1987).
- Tang, T.B. *Nature* **319**, 532 (1986).
- Bok, B.J., Cordwell, C.S. & Cromwell, R.H. in *Dark nebulae, globules and protostars* (ed. Lynds, B.T.) 33–35 (University of Arizona Press, Tucson, 1971).
- Spitzer, L. *Physical Processes in the Interstellar Medium* (Wiley, New York, 1978).
- Bruhweiler, F.C., Kondo, Y. & Sion, E.M. *Nature* **324**, 235 (1986).

unlike the proposed violent events⁶ affecting Sirius B. This could explain why the colour changes of Sirius were never reported as special events.

C. GRY

Laboratoire d'Astronomie Spatiale du CNRS,

Les Trois Lucs, F-13012 Marseille, France

J.M. BONNET-BIDAUD

Service d'Astrophysique,
CEN Saclay,
F-91191 Gif-sur-Yvette Cedex, France

Sex-ratio biased organisms

SIR—Hurst *et al.* reported in *News and Views*¹ the work of Stouthamer *et al.*² who showed evidence that a microorganism is the probable cause of sex-ratio biasing in the wasp genus *Trichogramma*. Treatment with some antibiotics and temperature above 30°C caused completely partheno-genetic wasps to adopt permanently bisexual reproduction. Stouthamer *et al.* did not isolate the supposed microorganism, but provided good circumstantial evidence for its existence. Hurst *et al.* did not mention the sex-ratio organism of the fruitfly *Drosophila* or of spiroplasmas. The reader could have been left with the impression that a biased sex ratio in offspring associated with microbial infection is indeed novel.

The novelty of the article by Stouthamer *et al.* stemmed rather from "the idea that the agents (microorganisms) might somehow cause uniparental reproduction", which, if it occurs, may well be novel. The non-mendelian nature of the sex-ratio phenomenon in *Drosophila* is well known³, and the microorganism responsible for it was first observed in 1961⁴. This organism is now known to belong to the bacterial genus *Spiroplasma* (class Mollicutes). Stouthamer *et al.* mention spiroplasmas indirectly, but failed to appreciate the many similarities their putative organism shared with them^{5,6}.

IAN HUMPHERY-SMITH

CLAUDE CHASTEL

Département de Microbiologie et Santé Publique,
Faculté de Médecine,
29285 Brest Cédex,
France

- Hurst, L.D., Godfray, H.C.J. & Harvey, P.H. *Nature* **346**, 510–511 (1990).
- Stouthamer, R., Luck, R.F. & Hamilton, W.D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2424–2427 (1990).
- Magni, G.E. *Nature* **172**, 81 (1953).
- Poulsen, D.F. & Sakaguchi, B. *Science* **133**, 1489–1490 (1961).
- McGarrity, G.J. & Williamson, D.L. in *The Mycoplasmas* (eds Whitcomb, R.F. & Tully, J.G.) Vol. 5 365–392 (Academic, New York, 1989).
- Werren, J.H. *J. theor. Biol.* **124**, 317–334 (1987).
- Chastel, C. & Humphery-Smith, I. *Adv. Dis. Vector Res.* **7**, 149–206 (1990).
- Williamson, D.L., Tully, J.G. & Whitcomb, R.F. in *The Mycoplasmas* (eds Whitcomb, R.F. & Tully, J.G.) Vol. 5 71–111 (Academic, New York, 1989).

Fractal dimension of co-citations

SIR—Like many natural phenomena, the growth of scientific knowledge seems to be cluster-like. On a spatial scale, scientific discoveries mainly 'cluster' around important research institutes. On a temporal scale, scientific discoveries often occur in relatively short periods of time as an important breakthrough makes new advances possible. Here I focus on a non-physical abstract structure in which pieces of scientific information are clustered according to specific aggregation rules. I discuss geometrical properties of co-citation clustering and in particular, the size distribution of these clusters in terms of fractal dimensions.

When a scientific paper cites two earlier papers, these latter papers are 'co-cited'. The strength of such a co-citation relation is determined by the number of citing papers having the above pair in their reference list. One of these co-cited papers can also form a co-citation pair with a third paper. In this way, clusters of (co-)cited papers emerge, and a 'map' of the citation field can be created¹. Looking at papers published in 1984, about six million cited papers are reduced to some 70,000 highly cited papers, and with these papers nearly 10,000 clusters are formed. These C1 clusters are used as input for a second-step clustering resulting in about 1,400 C2 clusters, each of them containing 2 to 60 clusters of the first (C1) cluster generation. The iteration procedure is then performed twice again, the 1,400 C2 clusters being input for about 180 C3 clusters, and these latter clusters being input for the final C4 clustering, yielding 21 C4 clusters. Each iteration reduces the number of clusters by an order of magnitude. In general, one may say that at the C1 level science is structured in terms of (small) research specialties, whereas at higher levels co-citation clusters become increasingly extensive in size, representing higher hierarchical structures such as sub-fields, fields and disciplines. 'Cluster size' refers to the number of citing papers involved.

In the figure, I present data on the size-rank distribution of C2 and C3 co-citation clusters. For the largest part of the ranking scale, over about two orders of magnitude, the distribution is a power law $g(r) = k r^{-\gamma}$ where $g(r)$ is the size (number of citing papers) of the cluster with rank r , k is a value determined empirically, and γ

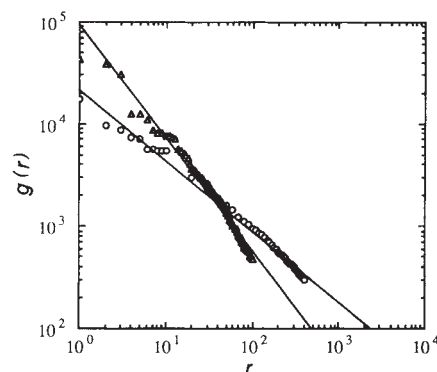
is the slope of the line. I found that $\gamma(\text{C3}) = 1.09$ and $\gamma(\text{C2}) = 0.71$, indicating a clear difference between the cluster systems. The next step is the conversion of the size-rank distribution function into the 'usual' size distribution function: the number of clusters with size g , $n(g)$, as a function of g

$$n(g) = -dr/dg \sim g^{-(1/\gamma + 1)} \quad (1)$$

Using fractal theory^{2,3}, the size distribution of fractal 'islands' (fragments of area A from fractal fragmentation) is

$$n(A) \sim A^{-(D/D_c + 1)} \quad (2)$$

where D is the fractal dimension describing the degree of fragmentation of the set of islands, and D_c is the fractal dimension



Size-rank distribution of C2 and C3 Co-citation clusters. Circles, distribution for the power-law part of the (total) 1,371 C2 clusters⁵; triangles, distribution for the power-law part of the (total) 179 C3 clusters (C3 data was collected using the online version of the Science Citation Index (SCISEARCH, at the host Deutsches Institut für Medizinische Dokumentation und Information, Cologne)).

of the 'coastline' of individual islands or fragments.

The 'island area' can be compared with size or 'volume' of the co-citation clusters, that is $g = A$. The citing papers constitute the cluster size and in geometrical terms may be considered as unit surface or unit volume elements. Comparing equations (1) and (2) gives $1/\gamma = D/D_c$. I am primarily interested in the fractal fragmentation dimension D of the co-citation cluster distribution. Assuming, to a first approximation, that the form of the co-citation clusters is regular (the clusters have a smooth 'coastline'), I put $D_c = 1$, and therefore $\gamma = 1/D$. This agrees with the Mandelbrot² generalization of the Zipf frequency distribution recently used to describe the size distribution of species in ecosystems⁴. For the C2 clusters, $\gamma(\text{C2}) = 0.71$, so that $D(\text{C2}) = 1.41$, and for the C3 clusters, $\gamma(\text{C3}) = 1.09$, which gives $D(\text{C3}) = 0.92$. The fractal dimension of C2 clusters is significantly higher than for C3 clusters. These results agree qualitatively with results for 'fractal landscapes' (which mimic those generated from brownian

motion⁵), in particular the relation between degree of fragmentation and the fractal dimensions of island size distributions². A striking characteristic of these fractal landscapes is that the higher the fractal dimension $1 \leq D \leq 2$, the more fragmentation of larger islands occurs.

What is the meaning of a fractal dimension of information space as represented by co-citation clustering? The size distribution of the co-citation clusters is a snapshot of a dynamical process, reflecting the presence of established fields and the emergence of new specialities. Like fractal distributions in ecological systems⁴ one may consider co-citation clusters as a representation of the ecosystem of scientists.

In this model, the structure of the ecosystem is strongly related to some optimal distribution of energy, mass and information. If co-citation clusters represent 'species of scientists', then the fractal cluster distribution gives a measure of the diversity of the research community, that is, the distribution of individuals among species owing to the optimization of flows of 'mass' (scientists, budgets) and information. For the small C2 and C3 clusters there is a deviation from the power-law behaviour and for such small clusters, the parameters that determine their structure and relations (such as flow of people, money and information) are much more subject to a random process, whereas for the larger clusters, the underlying dynamics follows particular patterns yielding fractal distributions. Fractal geometry is known to be closely related to the problem of describing the propagation of order in non-equilibrium systems. Therefore, the fractal model of co-citation clustering is an interesting starting point for further modelling of scientific ecosystems.

Very recently we determined the size distribution of the nearly 10,000 C1 clusters, the 'finest' fragmentation of science (in terms of co-citations). A fractal dimension of $D=2.0$ is found, which is in line with the findings for the C2 and C3 clusters.

A. F. J. VAN RAAN

Centre for Science and Technology
Studies,
University of Leiden,
Wassenaarseweg 52,
PO Box 9555,
2300 RB Leiden,
The Netherlands

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references. □

- Small, H. & Garfield, E. J. *Inf. Sci.* **11**, 147–159 (1985).
- Mandelbrot, B.B. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
- Voss, R.F., Laibowitz, R.B. & Alessandrini, E. I. in *Scaling Phenomena in Disordered Systems* (eds Pynn, R. & Skjeltorp, A.) NATO ASI Ser., **B133**, 279–288 (Plenum, London, 1985).
- Frontier, S. in *Developments in Numerical Ecology* (eds Legendre, P. & Legendre, L.) NATO ASI Ser., **G14**, 335–378 (Springer, Berlin, 1987).
- Weingart, P., Sehringer, R. & Winterhager, M. in *Handbook of Quantitative Studies of Science and Technology* (ed. Van Raan, A. F. J.) (North Holland-Elsevier, Amsterdam, 1988).

Mental simulation

Geoffrey Hinton

An Introduction to Neural Computing. By Igor Aleksander and Helen Morton. *Chapman and Hall*: 1990. Pp. 240. Pbk £15.95*.

RESEARCHERS in the fields of physics, computer science, electrical engineering, statistics, neuroscience, psychology and others are currently investigating the computational abilities of networks of non-linear neurons. Their scientific aim is to understand how the brain computes, and their technological goal, which is the main concern of this book, is to produce useful adaptive devices.

The central problem in the field of 'neural networks' is to find a learning algorithm that will adjust the connection strengths in the network so that the neural activities represent important underlying properties of the network's input. By associating its responses with these underlying properties, the network can behave in much more complex ways than a system that simply associates responses with the raw input. Much of the recent interest has been stimulated by the discovery of new learning procedures such as Kohonen's method of forming topographic maps or the back-propagation procedure described by Rumelhart *et al.* (*Nature* **323**, 533; 1986). These procedures are relatively painless to program on scientific workstations and work well for training small networks to solve small tasks, so curious researchers can get hooked in a few days.

For those entering the field of neural networks, there is an acute need for an authoritative textbook that explains the current models, points out their weaknesses, and relates them to their roots in established disciplines. Aleksander and Morton fail in this task because they assume a complete ignorance of calculus, linear algebra and statistics. They do a good job of explaining the basic models by using detailed examples, so their book will be valuable to those who want to get the flavour of the ideas without understanding the relevant mathematics, but they fail to provide a solid foundation that will transcend the specific models they describe. This is unfortunate, as any introductory book in this fast-moving field is inevitably out of date by the time it appears.

Aleksander and Morton cover most of the important models up until 1987 and their descriptions are fairly accurate. They do, however, make a few careless blunders, such as confusing the number of times

a neuron fires with the number of times it is given an opportunity to decide whether or not to fire (page 95). They also choose to describe the unsupervised version of the Boltzmann machine learning procedure instead of the more easily understood supervised version, and they seem unaware that the 'mean field' approximation used by physicists leads to an equivalent but much faster learning procedure that can be applied to analogue deterministic networks of the kind that Hopfield and Tank introduced for optimization tasks. The authors devote only 15 pages to applications of neural networks

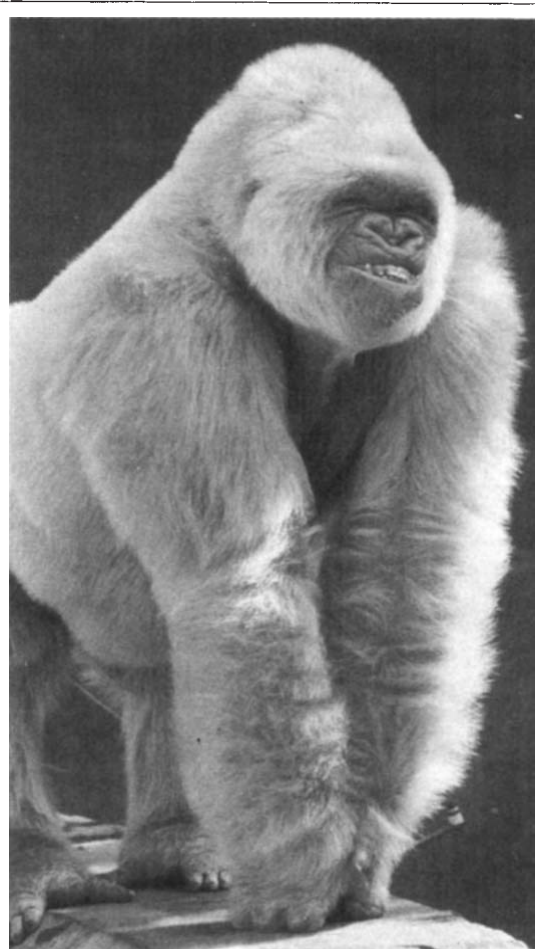
in speech, language and vision, so they inevitably omit some of the most significant practical successes such as the use of multilayer time-delay networks for phoneme recognition or the use of constrained multilayer networks for hand-written character recognition. Both these successes stem from ideas about how neural networks can achieve perceptual invariance so that their recognition is not affected by changes in the onset time of a phoneme or the position of a character.

Aleksander and Morton do not attempt to relate neural-network models to important ideas in other fields. For example, competitive learning can be viewed as a crude and inefficient approximation to a maximum likelihood fit of a mixture of gaussians, and back propagation can be viewed as a multilayer extension of logistic regression, or as a nonlinear generalization of methods used in optimal control. Hopfield networks, particularly the analogue version, are closely related to iterative techniques used within computer vision for processes such as stereo-fusion and surface interpolation.

A significant limitation of the book is that the authors do not seriously discuss the issue of generalization. There is no mention of theoretical work by Baum and Haussler (published in *Neural Computation* in 1989) which establishes that if a small network can be trained to give the correct response for a large num-

ber of examples, it will almost certainly generalize correctly to further examples. Nor do the authors mention that generalization can be improved by shrinking the weights, eliminating unnecessary connections or using a validation set to decide when to stop learning on the training set.

An even more significant omission is a discussion of the speed of learning. This is now the crucial issue in the practical application of neural networks because current algorithms such as back propagation do not scale up well to large networks. Greedy algorithms that add hidden units one at a time can greatly improve learning speed, but for really large networks it will probably be necessary to develop more modular systems in which individual modules can learn independently because they have local, internally generated goals that allow them to build appropriate representations of the underlying causes of the sensory input without requiring an external teacher. ►



Captivating smile — the albino gorilla, Snowflake, at the Barcelona Zoo attracts many fans. Traditional menageries were once the domain of the rich, where people went to be amused or amazed; modern zoos have been derided as soulless prisons. For many species zoos are now the last refuge and their only hope for survival. *Zoo. The Modern Ark* by Jake Page with photos by Franz Maier looks at zoos throughout history and examines their role in protecting endangered wildlife. Published by Facts On File, price is £18.95. □

* A software package, *Cortex*, is also available from Unistat Ltd, PO Box 383, London N6 5UP, UK. This package allows users to construct their own neural networks and is for use in conjunction with the book reviewed here. Price is £100.

Despite its limitations, this book is a helpful introduction for the non-mathematical novice. There is an accompanying simple simulator (see footnote) which, though much less extensive than the simulator that accompanies McClelland and Rumelhart's *Explorations in Parallel Distributed Processing* (MIT Press, 1989), is still useful in providing hands-on experience. □

Geoffrey Hinton is a fellow of the Canadian Institute for Advanced Research in the Department of Computer Science, University of Toronto, Toronto, Canada M5S 1A4.

Digestive clues

J.-J. Jaeger

Owls, Caves And Fossils. By Peter Andrews. *Natural History Museum Publications*: 1990. Pp. 231. £27.50.

WHEN palaeontologists or archaeologists discover a rich microvertebrate fauna, they are so concerned with its biochronological and palaeoenvironmental information content that they forget to ask questions about the origin of these small vertebrates' assemblages, the causes of their death and concentration, and alteration and/or transport. All these questions are the concern of a small but important field of palaeobiology called taphonomy. Peter Andrews in his new book draws attention to these fundamental points, which seem to be critical when information from microvertebrate faunas is used for palaeoenvironmental reconstructions. Of special interest are the numerous pictures taken with the scanning electron microscope of recent small bones showing different types of alterations resulting from processes such as weathering and digestion.

The book is organized into seven chapters, each concluding with a summary. The first three are devoted to the taphonomy of extant microvertebrates, principally represented here by rodents which have been eaten by owls. Diurnal birds of prey and mammalian predators of microvertebrates are much less important in this context as their diets are more diverse than that of owls and their digestion of bones and teeth more efficient; they therefore play only a minor role in the accumulation of microvertebrates in extant or fossil concentration sites.

Using information from original experiments performed over several years, detailed scanning electron microscope studies of bone and teeth surfaces, and precise calculation of proportions and indices, Andrews discusses in detail the science of owl pellets and its contribution to the fossil record. The results allow one to distinguish between, for instance,

Barred owl — the remains of its prey could turn out to be useful palaeoclimate indicators.

alteration of bone and teeth caused by soil-forming processes or by digestive etching: each has a characteristic pattern of surface destruction. Also, detailed study of the relative proportions of different bones and the degrees of breakage and digestion in specific accumulations indicates the possibility of distinguishing concentrations due to diurnal birds of prey and mammalian predators from those of owls, and even to identify the species of owl involved in the bone accumulation.

Nevertheless, the method is still at a descriptive stage and, with so many data, I wonder whether multivariate statistical analysis would not have been a more efficient tool than single ratios or indices. Moreover, it is stated that predators in the fossil assemblage are least likely to have been involved in its accumulation, but this is true only in the case of short-duration accumulations. When the duration increases, then the probability of finding some predator's bones mixed with the bones of its prey increases. For instance, in Italy, fossil giant rodents have been found in fissure fillings in association with bones of a giant owl of Pliocene age, an example of co-evolution of size increase between predator and prey. Because of the interest of the proposed method, it is a little unfortunate that no mention is made of the karst fissure filling or cave fillings older than the Pleistocene. Some have a surprising antiquity, such as the oldest European Palaeocene mammal locality of Walbeck in eastern Germany.

The second part of the book is devoted to a case study; the small mammalian fauna of the Westbury-Sub-Mendip cave in Somerset, southwest England. Andrews applies his method to several levels containing microvertebrate concentrations of

middle Pleistocene age, and demonstrates that the identification of predators is possible only when there has been almost no large-scale transport or post-accumulation trampling. Andrews also demonstrates that severe breakage and alteration can occur during transport, thus making it almost impossible, in most cases, to identify the initial predator. Nevertheless, a precise identification can be proposed for most concentrations, and Andrews discusses the palaeoecological significance of these microfaunas. He shows that the most accurate method for palaeoecological analysis is provided by the use of the taxonomic habitat index, a semi-quantitative method.

This book has therefore to be considered as the state of the art for work in the field of small-mammal taphonomy. Very few detailed papers have been devoted to this subject, although many palaeontologists and archaeologists have individually accumulated empirical experience in the field. These people will therefore welcome this book because not only has Andrews collected and discussed all the available evidence but he has also tried to elaborate a semi-quantitative method for identification of predators based on detailed qualitative and quantitative study of the bones of prey species.

This method could be of particular use in the quantitative reconstruction of past climates. Microvertebrates could turn out to be even more precise palaeoclimate indicators than pollen or land snails, as their taphonomy relies on more detailed analyses such as those discussed here. □

J.-J. Jaeger is at the Institut des Sciences de l'Évolution, Place Eugène-Bataillon, 34095 Montpellier, Cedex 05, France.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Higher learning

Richard Davenport-Hines

Medieval and Early Renaissance Medicine. An Introduction to Knowledge and Practice.

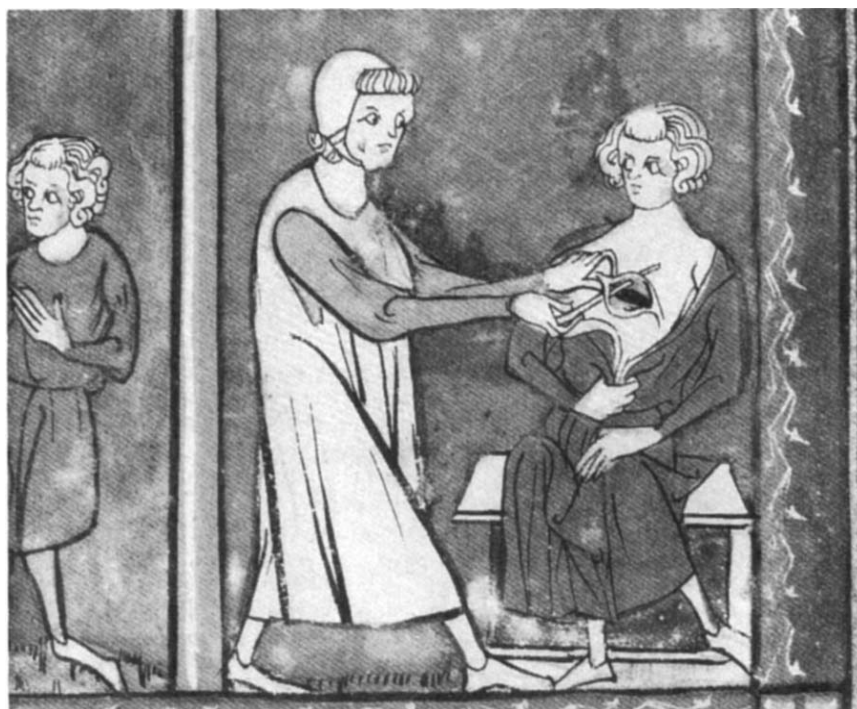
By Nancy G. Siraisi. Chicago University Press: 1990. Pp. 250. Hbk \$37.50, £29.95; pbk \$10.95, £8.75.

THE realities of health and disease, the needs of patients and their experiences at the hands of medical practitioners cannot fail to make interesting reading. Physicians are in intimate contact with their patients at the great human crises of birth and death, and the story of their professional lives is always rich in human significance. Yet given the promising subject of her book, Nancy Siraisi's study of mediaeval medical practices in western Europe is strangely disappointing. The tone of her monograph is unnecessarily desiccated and, except in her more lively chapter on surgical techniques, gives little sense of the human lives which she is chronicling.

Siraisi begins by tracing the dominant influences of ptolemaic astronomy, aristotelian logic, epistemology, cosmology and physics on the conceptualization and practice of medicine in the four centuries before 1500. She traces the ancient and Islamic antecedents of west European

medical ideas, showing the continuing importance of hippocratic authors and the galenic system. The gradual multiplication of medical books and their accumulation in monasteries and later universities steadily increased the theoretical, systematic and learned elements in medicine until a rich, specialised literature had developed. Beginning in Salerno, medicine became a part of university curricula

about physicians' status and emoluments. The criticism of the better-educated mediaeval physicians that they are pursuing for mercenary motives subjects that should be studied out of charity or for love of pure knowledge will be familiar to many contemporary readers. The ways in which christianization resulted in new centres of healing (with monks turning some monasteries into medical centres),



From *Medieval and Early Renaissance Medicine*

A stitch in time — in order clearly to illustrate the suturing procedure, the wound in this fourteenth-century miniature has been greatly exaggerated.

and, as shown by the life of the fifteenth-century literary humanist and medical practitioner Michele Savonarola, the subject was accepted as one of the higher branches of learning. Siraisi describes the extent of physiological and anatomical knowledge and surgical techniques, and other treatments such as phlebotomy, blood-letting and herbalism. She also shows the primacy in mediaeval medical theory of belief in the four humours, and in the use of concept of complexion to identify the different elements of the human body — a set of beliefs which were finally shaken by the epidemic of syphilis in the 1490s.

Some of the most successful passages of her book concern questions of professional identity. The gradual professionalization of medical practice, the contrasts between lettered practitioners in rich, urban areas and lowly empirics in poor, rural districts, the ways in which medical specialization became a mark of low status in practitioners, the employment by mediaeval municipalities of distinguished medical men, the persistent sub-cultures of women surgeons and of Jewish physicians — these historical developments are all discussed, in ways which recall many familiar twentieth-century controversies

and the interplay between ideas of healing and faith, or evil and disease, are solidly but not searchingly described.

Although Siraisi's book is offered as a resumé of current historical understanding and makes little pretence of offering ambitious or original insights for specialist scholars, some sections are abstruse and will make taxing reading for the non-specialist. There is a tendency for truisms to be enunciated with ponderous importance as if they are astonishing new insights, as for example when readers are allowed to share Siraisi's dazzling realization that a talent for prognosis "was an important aspect of the physician's skill". Some compensation for the laboured prose is given by many excellent black and white illustrations of medical techniques taken from mediaeval or Renaissance treatises or manuscripts. These are informatively captioned, and provide a fascinating sense of medical iconography in a pre-literate society. Overall, *Medieval and Early Renaissance Medicine* is a reliable but very far from inspired general introduction to its subject. □

Richard Davenport-Hines, 51 Elsham Road, London W14 8HD, UK is a freelance medical historian.

New in paperback

■ *Lemurs of the Lost World*, by Jane Wilson, describes the zoological wealth of the caves and forests of northern Madagascar. Published by Harrap, price is £5.95.

■ New in paperback is *Hamiltonian Systems: Chaos and Quantization* by Alfredo M. Ozorio de Almeida. First published as a Cambridge Monograph on Mathematical Physics in 1988, the paperback costs £15, \$24.95.

■ *The Crisis of Life on Earth* is a science journalist's account of "how natural forces and human folly have combined to bring our planet to the edge of an apocalypse". By Tim Radford, formerly science editor of the *Guardian* newspaper, publisher is Thorsons, price £6.99.

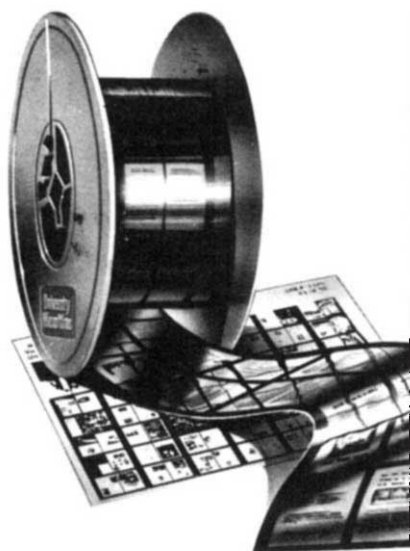
■ The latest in the IRL Press 'Practical Approach' series is *Receptor Biochemistry*, edited by E. C. Hulme. The book is the first in a set of three covering receptor-effector coupling and neurotransmitter/hormone receptor-ligand interactions. Publisher is Oxford University Press, price is £25. (Also available spiral bound at £35.)

■ Penguin books has just published *Galileo: Heretic* by Pietro Redondi. The first English edition was published by Princeton University Press in 1987, and the book was reviewed in *Nature* that year (330, 617). The paperback price is £7.99.

■ *Biotechnology in the Food Industry* by M. P. Tombs is published by the Open University Press in association with the Institute of Biology. The book is for undergraduates and postgraduates, and the author aims to point out the benefits and the drawbacks of this controversial subject. Price £14.99. (Also published in hardback at £35.) □

nature

is available in microform.



University Microfilms International reproduces this publication in microform: microfiche and 16mm or 35mm film. For information about this publication or any of the more than 13,000 titles we offer, complete and mail the coupon to: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106. Call us toll-free for an immediate response: 800-521-3044. Or call collect in Michigan, Alaska and Hawaii: 313-761-4700.

☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

City _____

State _____ Zip _____

Phone () _____

**University
Microfilms
International**

Uncertain gift

Brian Pippard

Chaotic Dynamics. An Introduction. By G. L. Baker and J. P. Gollub. Cambridge University Press: 1990. Pp 182. Hbk £25, \$49.50; pbk £9.95, \$17.95.

CHAOS must be counted among the gifts the computer has brought to science, for it is only after immensely long computations of the development of nonlinear dynamical systems that the pervasiveness and complexity of chaos have come to be recognized. It could be argued that its study is hardly a branch of science, as the most characteristic property of a chaotic system — its extraordinary sensitivity to minute perturbations — usually permits only the coarse features to be demonstrated in practice. Mathematicians seek their prey among the fine details, and they have the promise of many seasons of good hunting. But science cannot remain aloof, for chaos, revealing as it does the limits to predictability, must ultimately affect the philosophy of science in general, as well as the private philosophies of individual scientists.

The authors of this introduction to the subject have sought to bridge the chasm between the modest mathematical skills of a science student and the rather appalling complexity of most writings about chaos, as well as the *art nouveau* of fractal geometry. With the help of copious computer drawings and elementary exercises they have succeeded in conveying basic concepts in a technically simple manner. The occasional lapses, such as the use in a single diagram of the symbol f to mean both a function of time and a frequency of oscillation, are only momentary distractions in an otherwise easily flowing argument which should not discourage even a timid novice.

Whether the novice will be actively encouraged is another matter, for I am not persuaded that the problem has been attacked from the best angle. Much of the development takes as exemplar the hard-driven pendulum. Excellent as an easily visualized model, the pendulum suffers the disadvantage of being governed by a differential equation that cannot be solved in terms of elementary functions. Thus one must rely on point-by-point numerical integration which, even with a large computer, makes the collection of data on the long-term behaviour a slow process. Moreover, a reader who knows how sensitive a chaotic system is to minute disturbances must wonder about the adequacy of the Runge-Kutta program to avoid artefacts; and nothing is said to allay his doubts. It would surely have been better to choose a model which allows one to proceed analytically, and therefore

enormously faster, from one point on a Poincaré section to the next. Balls bouncing on oscillating plates, and other impact oscillators, have this simplicity and are just as easily visualized; data generated with even a small desk computer verify some of the theoretical predictions more convincingly than seems to have been possible with the pendulum.

Although so much emphasis is placed on one special system, it does provide opportunities to illustrate many basic ideas — bifurcation, Lyapunov exponent, winding number, fractal dimension. The reader, however, may wonder why anyone bothers to introduce them as the all-too-brief final chapter, devoted to physical applications, gives (apart from bifurcation) no convincing example of their use. A mathematician can doubtless love these concepts for their own sake, but will a scientist find an incentive to take things further? Whoever is fired to do so must harbour no illusions; this gentle introduction gives scant warning of the arduous task to be undertaken by any who desire to achieve professional competence. □

Brian Pippard is at the Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK.



"Nothing has been so destructive to indigenous peoples as what we call progress. Mines, dams, roads, . . . and other expressions of 'economic development' have forced indigenous peoples from lands they have occupied for centuries. . . ." *The Gaia Atlas of First Peoples* by Julian Berger examines the way of life of the indigenous peoples of the world and the ecological, economic and political threats which they face. Publisher is Robertson McCarta, price is £13.95 (hbk), £8.95 (pbk). □

Molecular dynamics simulations in biology

Martin Karplus & Gregory A. Petsko

Molecular dynamics—the science of simulating the motions of a system of particles—applied to biological macromolecules gives the fluctuations in the relative positions of the atoms in a protein or in DNA as a function of time. Knowledge of these motions provides insights into biological phenomena such as the role of flexibility in ligand binding and the rapid solvation of the electron transfer state in photosynthesis. Molecular dynamics is also being used to determine protein structures from NMR, to refine protein X-ray crystal structures faster from poorer starting models, and to calculate the free energy changes resulting from mutations in proteins.

SCIENTIFIC theory has as its ultimate goal the understanding and prediction of natural phenomena. In physics and chemistry, theory and experiment have worked well together to interpret the results of measurements and to provide predictions that aid in their design. Biology has been slower to embrace theoretical approaches, in part because of a certain scepticism as to the validity of the assumptions and methods that are applicable to large and complex systems, such as the macromolecules of biological interest, not to mention organelles or cells. The most successful theories make use of a combination of analytical and computational approaches. They have given rise to 'theoretical biophysics', a rapidly growing field that is having a significant impact on many areas of biological research.

Supercomputer availability has contributed to the development of this field, but the main reason for its growth is the introduction and application of molecular dynamics simulation methods^{1,2}. Biological macromolecules are inherently dynamic systems, and it was the demonstration that calculations based on simple physical models could give new, and testable, insights into their internal motions that convinced many biochemists that biophysical theory had become relevant. Dynamical simulations have now been extended to include not only molecular motion, but also the refinement of macromolecular structures based on X-ray and NMR data and the sampling of configuration space required for the evaluation of free energy changes induced by mutations and other perturbations.

Macromolecular motions cover a large range in magnitude (from hundredths of an ångström to tens of ångströms), and an enormous range of time (from sub-picoseconds to seconds and even longer). Some of these motions are known to have a functional role. Oxygen cannot reach the binding site in myoglobin or haemoglobin unless the protein structure fluctuates to open a transient pathway. Hexokinase transfers a phosphate from ATP to glucose much faster than it transfers one to water because the binding of glucose induces a large conformational change in the protein that brings two domains together and creates a protected catalytic site. The domain closure is slow, but it would not be possible without the numerous rapid atomic motions that allow groups of atoms to slide past each other despite their repulsive interactions. There are many other examples in biology in which dynamics is involved, either as the essential element of the phenomenon or as a secondary element (that is, the timescale of the motion is not important). Molecular dynamics simulations are important for understanding these motions, particularly because detailed experimental studies are often very difficult.

This review focuses on the methodology of molecular dynamics and its applications to biological problems. We begin with an overview of the methodology, including some remarks on its experimental validation, and continue with an account of some important applications. We focus on proteins in this review, although similar but less extensive studies have been made for nucleic acids, oligosaccharides and lipids. We close with a discussion of the present limitations of molecular

dynamics simulations and some comments on likely future developments.

What is molecular dynamics?

Molecular dynamics is the science of simulating the motions of a system of particles. It has been applied to systems as small as an atom and a diatomic molecule undergoing a chemical reaction, and as large as a galaxy³. In all cases, the essential elements for a molecular dynamics simulation are a knowledge of the interaction potential for the particles, from which the forces can be calculated, and of the equations of motion governing the dynamics of the particles. The interaction potential may vary from the simple gravitational interaction between stars to the complex many-body forces between atoms and molecules. Classical newtonian equations of motion are adequate for many systems, including the biomolecules of primary concern here, but for some problems (such as reactions involving tunnelling) quantum corrections are important, and for others (such as galaxy evolution) relativistic effects may have to be included.

Two attributes of molecular dynamics simulations have played an essential part in their explosive development and wide range of applications. Simulations provide individual particle motions as a function of time so they can be probed far more easily than experiments to answer detailed questions about the properties of a system. Further, although the potential used in a simulation is approximate, it is completely under the user's control, so that by removing or altering specific contributions their role in determining a given property can be examined. 'Computer alchemy'—changing the potential from that representing one system to another during a simulation—is a powerful tool for calculating free energy differences.

Potential energy function

Molecular dynamics simulations begin with a knowledge of the energy of the system as a function of the atomic coordinates. The potential energy surface determines the relative stabilities of the different possible stable or metastable structures. The forces acting on the atoms of the system, which are related to the first derivatives of the potential with respect to the atom positions, can be used to calculate the dynamic behaviour of the system by solving Newton's equations of motion for the atoms as a function of time.

Although quantum mechanical calculations can yield potential surfaces for small molecules, only empirical energy functions can provide this information for proteins and their environment. As at ordinary temperatures most motions leave the bond lengths and angles of polypeptide chains near their equilibrium values, the accuracy of the energy function representation of bonding is comparable to that of vibrational analyses of small molecules. In globular proteins, contacts between non-bonded atoms also make a large contribution to the potential energy of the folded or native structure. The success of hard-sphere non-bonded radii in conformational studies suggests that relatively simple functions may adequately describe them.

The energy functions used for proteins are generally composed of bonding terms representing bond lengths, bond angles, and torsional angles, and non-bonding terms consisting of van der Waals interactions and electrostatic contributions. One widely used expression⁴ is:

$$E(\mathbf{R}) = \frac{1}{2} \sum_{\text{bonds}} K_b (b - b_0)^2 + \frac{1}{2} \sum_{\text{bond angles}} K_\Theta (\Theta - \Theta_0)^2 + \frac{1}{2} \sum_{\text{torsional}} K_\phi [1 + \cos(n\phi - \delta)] + \sum_{\text{nb pairs}} \left(\frac{A}{r^{12}} - \frac{B}{r^6} + \frac{q_1 q_2}{Dr} \right) \quad (1)$$

The energy, E , is a function of the Cartesian coordinate set, $\mathbf{R}$, specifying the positions of all the atoms, from which are calculated the internal coordinates for bond lengths (b), bond angles (Θ), dihedral angles (ϕ) and interparticle distances (r).

The first term in equation (1) represents instantaneous displacements from the ideal bond length, b_0 , by a Hooke's law (harmonic) potential. Such a harmonic potential is the first approximation to the energy of a bond as a function of its length. The bond force constant K_b determines the flexibility of the bond and can be evaluated from infrared stretching frequencies or quantum mechanical calculations. Ideal bond lengths can be inferred from high resolution, low temperature crystal structures or microwave spectroscopy data. The energy associated with alteration of bond angles given by the second term in equation (1) is also represented by a harmonic potential. For rotations about bonds, torsion angle potential functions given by the third term in equation (1) are used. This potential is assumed to be periodic and modelled by a cosine or sum over cosine functions. The final term in equation (1) represents the contribution of non-bonded interactions and has three parts: a repulsive term preventing atoms from interpenetrating at very short distances; an attractive term accounting for the London dispersion forces between atoms; and an electrostatic term that is attractive or repulsive depending on whether the charges q_1 and q_2 are of opposite or the same sign. The first two non-bonded terms combine to give the familiar Lennard-Jones 6-12 potential, which has a minimum at an interatomic separation equal to the sum of the van der Waals radii of the atoms; parameters A and B depend on the atoms involved and have been determined by a variety of methods, including non-bonding distances in crystals and gas-phase scattering measurements.

Electrostatic interactions between pairs of atoms are represented by a Coulomb potential with D the effective dielectric function for the medium and r the distance between the two charges. Use of atomic partial charges avoids the need for a separate term to represent the hydrogen bond interaction; that is, when the positive hydrogen attached to an electronegative atom comes within van der Waals distance of a negative acceptor atom, the Coulomb attraction adds to the Lennard-Jones potential and results in a hydrogen bond.

The usefulness of empirical energy functions depends on the extent to which the parameters determined for equation (1) by the study of model systems, such as amino acids, can be employed for macromolecules, such as proteins. Evidence from a number of comparisons suggest that this transferability condition is satisfied in many applications.

Simulation methods

Given a potential energy function there are several approaches to the dynamics. The most exact and detailed information is provided by molecular dynamics simulations in which Newton's equations of motion are solved for the atoms of the system and any surrounding solvent. For a simple homogeneous system, such as a box of water molecules with periodic boundary conditions, average structural and dynamic properties can be determined in simulations of only a few picoseconds⁵. Inhomogeneous systems such as proteins require considerably longer simulations. Modern computers allow simulations of up to nanoseconds, long enough to completely characterize the librations of small groups in a protein and to determine the dominant contributions to atomic fluctuations¹.

To begin a dynamic simulation an initial set of atomic coordinates and velocities is required. The coordinates can be obtained from X-ray crystallographic or NMR structure data, or by model-building (based on the structure of a homologous protein, for example). Given a set of coordinates, a preliminary calculation serves to equilibrate the system. The structure is first refined using an iterative minimization algorithm to relieve local stresses due to overlaps of non-bonded atoms, bond length distortions, and so on. Next, atoms are assigned velocities (v) taken at random from a maxwellian distribution for a low temperature, and a simulation is performed for a few picoseconds. This is done by finding the acceleration a_i of atom i from Newton's law $F_i = m_i a_i$. (F_i , the force on the atom, is computed from the derivatives of equation (1) with respect to the position; m_i is the atomic mass), and introducing it into the equation for the

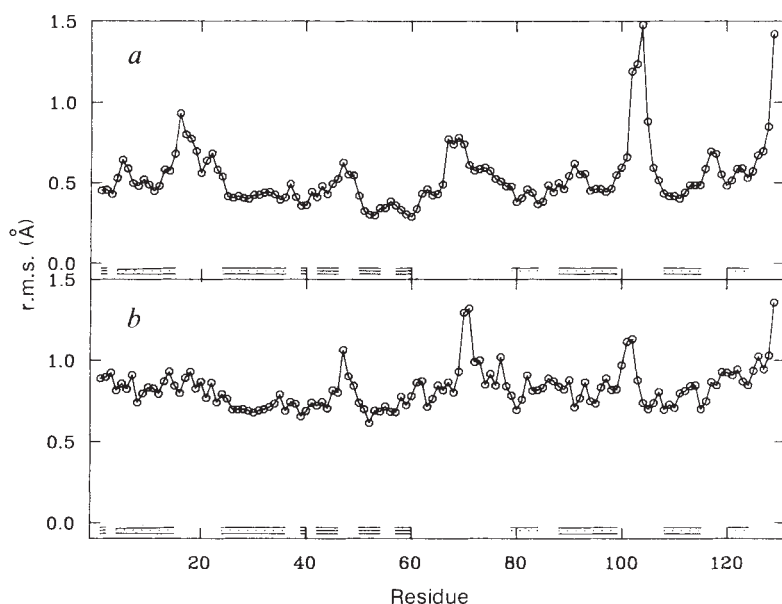


FIG. 1 Calculated and experimental r.m.s. fluctuations of lysozyme. Backbone averages are shown as a function of residue number. They were obtained *a*, from a molecular dynamics simulation, and *b*, from X-ray temperature factors without correcting for disorder contributions (from ref. 28).

position $\mathbf{r}_i$ at time $t + \Delta t$, given $\mathbf{r}_i$ at time t :

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i \Delta t + \frac{1}{2} \mathbf{a}_i (\Delta t)^2 + \dots \quad (2)$$

The equilibration is continued by alternating new velocity assignments, chosen from maxwellian distributions for temperatures that are successively increased to some chosen value, with intervals of dynamical relaxation. The temperature T of the system is measured by the mean kinetic energy,

$$\frac{1}{2} \sum_{i=1}^N m_i \langle v_i^2 \rangle = \frac{3}{2} N k_B T \quad (3)$$

where N is the number of atoms in the system, $\langle v_i^2 \rangle$ is the average velocity squared of the i th atom and k_B is the Boltzmann constant. The equilibration period is considered finished when the temperature is stable for longer than about 10 ps, the atomic momenta obey a maxwellian distribution and different regions of the protein have the same average temperature.

Continued integration of the equations of motion after equilibration generates the coordinates and velocities of the atoms as a function of time. With modern computers, equation (2) can be solved simultaneously for all atoms in a solvated macromolecule by use of a number of standard programs⁴. The quantity Δt must be very small so that the potential energy does not change too much during each time step; a satisfactory value for Δt is one femtosecond (10^{-15} s), which corresponds to about 30 steps for one vibration of a carbon-carbon bond. Even on a supercomputer such as the Cray X-MP a single time-step in the simulation of myoglobin (which has roughly 1,500 atoms) takes 0.2 s of CPU time, so that the 100,000 time-steps of a 100-ps simulation would take about 6 h of supercomputer time. The time increases roughly linearly with the number of atoms in the system.

For large systems the required computer time can be almost prohibitive. Nowhere is this problem more manifest than in the treatment of the solvent. Proteins and nucleic acids have evolved to function in an aqueous environment; large numbers of water molecules are tightly bound to their surfaces, forming a first hydration shell that is an integral part of the structure and should be included in the simulation. That can add hundreds of atoms to the computation. Bulk solvent is an even more serious problem—if not included the simulation is ‘*in vacuo*’, which may not affect interior atoms but is likely to have a serious effect on the conformations and motions of surface residues, especially those in exposed loops. Solvent must be included in simulations concerned with the quantitative analysis of thermodynamic properties, particularly of those portions of the protein at the solvent interface. For myoglobin, a satisfactory solvated simulation system includes about 1,000 water molecules (3,000 atoms) with a corresponding increase in computer time.

One strategy commonly used for simplifying the treatment of such large systems is to restrict the computation to a subset of atoms close to the site of interest (such as the active site of an enzyme). The atoms in a surrounding ‘buffer’ region are treated as stochastic particles, and all other atoms in the system are eliminated; their effect is approximated by boundary forces, frictional drag and stochastic terms acting on atoms in the buffer-region¹. This stochastic boundary molecular dynamics approach can reduce the computing time by a factor of 10 to 100, relative to a fully solvated simulation.

Experimental verification

For homogeneous systems, such as a box of water molecules, detailed verification of the structural and dynamic behaviour predicted by simulations has been possible by comparison with experiment⁵, but this is more difficult for the inhomogeneous systems of biology. Nevertheless, the approximate agreement between the average structure obtained in the first protein simulation⁶, that of the bovine pancreatic trypsin inhibitor (BPTI), and the structure determined by X-ray crystallography suggested

that such simulations were stable and that meaningful results could be obtained.

The fast motions sampled by the simulations leave their trace in X-ray diffraction data as they cause atoms to occupy more space on average than in a rigid molecule, smearing out the electron density, the quantity determined in crystallographic studies. If a gaussian model is used to fit the spread of density around the average position of each atom, the width of the gaussian can be expressed in terms of a ‘Debye-Waller factor’, or ‘isotropic temperature factor’, B , of the form:

$$\langle \Delta r_i^2 \rangle = 3 B_i / 8 \pi^2 \quad (4)$$

where $\langle \Delta r_i^2 \rangle$ is the mean-square fluctuation of atom i . Comparisons of $\langle \Delta r_i^2 \rangle$ values from molecular dynamics simulations and measured B factors show good qualitative agreement (Fig. 1). Larger deviations occur in regions of the structure where molecules make contact with neighbours in the crystal lattice: contact residues have systematically low experimental B factors. Surface side chains protruding from the protein also show a less good fit, but the discrepancy is reduced if the simulation is carried out in the presence of solvent and a crystalline environment⁷.

Internal motions of proteins can be studied also by NMR, which unlike X-ray diffraction is sensitive to the timescale as well as the magnitude of the motions⁸. NMR parameters such as nuclear spin-spin coupling and chemical shift are affected by molecular motion (Fig. 2). If interconversion of different

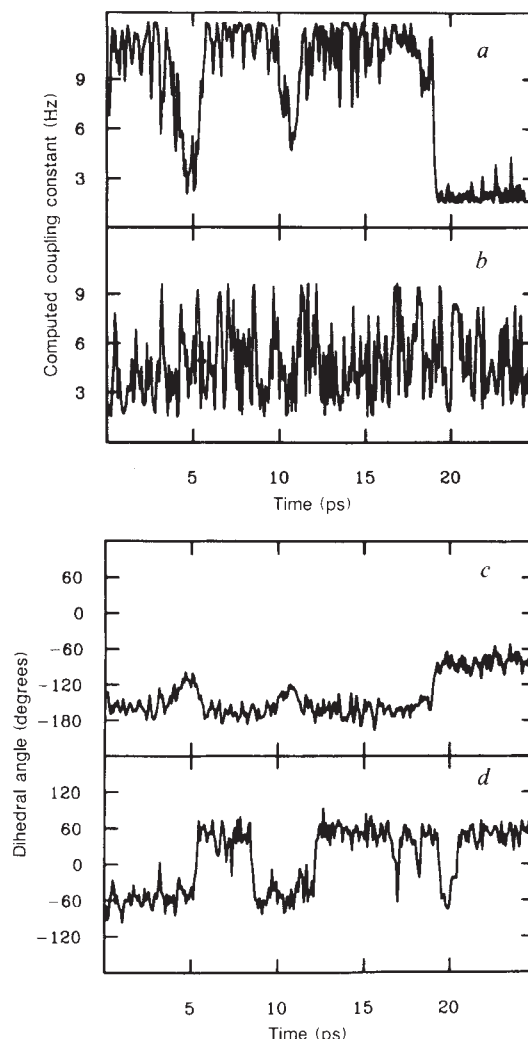


FIG. 2 Fluctuations of three-bond proton coupling constants and associated dihedral angles during a molecular dynamics simulations of BPTI. a, Glu 7 $^3J_{\text{BY}}$; b, Arg 42 $^3J_{\text{BY}}$; c, χ_2 of Glu 7; d, χ_2 of Asp 42 (from ref. 29).

local conformations is rapid on the NMR timescale, average values are obtained. For motions on the NMR timescale, transition rates can be determined directly by NMR. For example, aromatic residues are sometimes observed to 'flip' by NMR (on the timescale of milliseconds) and in long or activated dynamics simulations. That such ring flips occur, even in the tightly packed interior of proteins, came as a surprise to X-ray crystallographers and provided an early example of the importance of packing defects and correlated motions^{8,9}.

The ring flips constitute a rare event that is measurable because the result of flips (averaging of certain chemical shifts) has a significant effect on the NMR spectrum. One can measure even less frequent events if they are essential to the phenomenon under consideration. A case in point is hydrogen exchange, which may occur on a timescale of minutes or hours or even longer. Some hydrogens in the interior of proteins apparently can exchange only if a relatively large fluctuation takes place. Even though this happens only very rarely for a protein under physiological conditions, the rate can be measured when it controls a measurable phenomenon such as hydrogen exchange. Such complex events are still outside the range of what is directly accessible to molecular dynamics simulations.

Macromolecular motions

The picture that has emerged for the internal motions of proteins (and nucleic acids) is that the atoms are fluctuating on the picosecond timescale so that the instantaneous structures differ significantly from the time-average. This view has now become part of the dogma of biology and is often introduced in a descriptive fashion to explain a wide range of phenomena, even when there is no evidence that motions play a part.

During simulations the protein samples many conformations close to but not identical with the X-ray crystal structure. Atoms in the interior of proteins move less than atoms on the surface. Most atoms move anisotropically, and the directions of motion are determined more by non-bonded interactions with neighbouring groups than by the nature of the covalent bonds. Root-mean-square (r.m.s.) atomic displacements of 0.5–1 Å are not uncommon in the protein interior, and even larger ones occur for surface residues. The motions are somewhat anharmonic at 300 K, usually due to side chains 'hopping' between two or more alternative minima. A residue executes many small fluctuations about the minimum energy position before hopping to another minimum. As expected for motion involving a barrier the transition is very rapid, so that the side chain spends most of its time

1. Myoglobin and haemoglobin. Myoglobin and haemoglobin bind oxygen or carbon monoxide reversibly, yet the haem group is so deeply buried within the protein that there is no 'path' from outside the molecule to the ligand binding site. In their original papers, Perutz and Kendrew recognized that these proteins must 'breathe' to open channels connecting the surface and haem pocket. In the X-ray structure the energy barriers to ligand migration are high, with a half-time of approximately 10^{50} years required for entrance or exit of a ligand in the rigid protein. When the side chains surrounding the haem pocket are allowed to fluctuate, however, the barriers are substantially reduced and low energy pathways appear. One of these requires partial rotation of the distal histidine, His 64, and takes the ligand past one of the haem propionic acid side chains and out of the molecule in the vicinity of Arg 45, a pathway for which there is some experimental support from structural studies.

The mechanism can be examined in more detail by a molecular dynamics simulation of a ligand migrating into or out of the fluctuating protein. This was done by breaking the ligand-iron bond and giving the ligand a high initial velocity. In the absence of a bond to the iron, the non-bonding van der Waals repulsions prevent rebinding and the ligand starts to explore the distal pocket. Eventually some of the surrounding atoms move out of the way enough to allow the ligand to escape from the pocket and migrate through the protein until it finally reaches the outside.

To delineate the full range of possible pathways would seem to require many ligand plus protein simulations. A simplification is to run one simulation with hundreds of ligand molecules in the haem pocket, each with a different initial velocity¹⁷. Interaction terms between the ligand molecules are eliminated so that they do not 'feel' each other's presence, but each ligand molecule has the full interaction with the surrounding protein. As it is the dynamics of the protein that takes most time in the simulation, calculation for one myoglobin and many ligand molecules is very efficient. Such a 100-ps simulation of myoglobin with 60 carbon monoxide molecules revealed five principal routes for entrance and exit of a ligand. In all cases the ligand spent a considerable time rattling around inside one or another of the many atom-sized internal cavities known to exist in the protein before jumping over a barrier to a new position, usually another cavity or, finally, the protein exterior (Fig. 5).

in the minima, and it is these positions that are observed in the X-ray structure, as for the atoms of aromatic residues involved in ring flips described above. Below 160 K, simulations indicate that the dynamics are essentially harmonic. These features are confirmed by the temperature dependence of the *B* factors determined by X-ray crystallography and inelastic neutron scattering^{10–12}.

Two general models have been considered for the internal motions of proteins, and these are illustrated in Fig. 3*a*. In one model the fluctuations are assumed to occur within a single multidimensional well that is harmonic or quasiharmonic as a limiting case. The other model assumes there are multiple minima or substates, with the internal motions corresponding to a superposition of oscillations within wells and transitions between them. Experimental data have been interpreted according to both models, and it has proved difficult to distinguish between them. The two models can be distinguished by comparing structures sampled in a simulation, and determining whether they correspond to the same or different minima¹³. This is done by minimizing the coordinate sets obtained from the simulation (see Fig. 3*b*). In a 300 ps simulation of myoglobin at 300 K, different minima were accessed every few tenths of a picosecond, consistent with the model in which there are many structurally similar substates in the neighbourhood of the absolute minimum. The large number of minima sampled by the room temperature simulation suggests that the effective barriers separating them are low. As the energy differences between many of the wells are small, individual molecules may be trapped in metastable states at low temperature. This could lead to a glass-like substance.

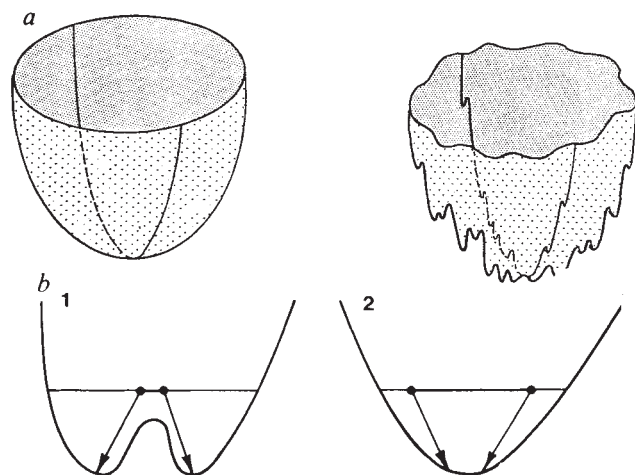


FIG. 3 Model potentials for protein motions. *a*, Two-dimensional representation of harmonic potential (left) and multimimum (substate) potential (right). *b*, representation of the r.m.s. difference criterion for different minima: 1, r.m.s. after the minimization is larger than the initial r.m.s., implying that the two conformations correspond to different minima; 2, r.m.s. after the minimization is smaller than the initial r.m.s., implying that the two conformations correspond to the same minimum (from ref. 13).

The backbone structural differences between the minima correspond to loop displacements coupled to changes in the relative orientation of helices which individually behave as nearly rigid bodies. The helix displacements are made possible by the loop rearrangements and by side-chain reorientations at helix-helix contacts. Specific loop motions are initiated by side-chain transitions in the helix contacts, main chain dihedral angle transitions of the loops themselves, or a combination of the two.

The dynamical results for the helix motions can be compared with structural data. The maximum dynamical displacements in the simulation are larger than those observed in different X-ray structures of the same protein and are of the same order as the differences found in the known X-ray structures of various globins. A comparison of the globin structures suggested that the observed range of helix packings is achieved primarily by changes in side chain volumes resulting from amino acid substitutions^{14,15}. In the dynamics, it is the correlated motions of side chains that are in contact, plus the rearrangements of loops, that make possible the observed helix fluctuations. As more than one set of side-chain orientations is consistent with a given set of helix positions, the known globin crystal structures represent only a very small subset of the possible local minima. Overall, the range of conformations sampled by a single myoglobin trajectory at 300 K is similar to that found in the evolutionary variation among structures of the globin series (Fig. 4). This 'molecular plasticity' is likely to have played a part in the evolution of protein sequences.

Functionally relevant motions

It would not be surprising if internal motions had been subjected to selective pressure during evolution. Just as structure is selected on the basis of function, there could be selection for certain internal motions, a consequence of the structure, if they had specific functional roles. Two examples of functionally relevant motions, one in globins and the other in photosynthetic reaction centres, are described in Boxes 1 and 2.

The myoglobin simulations (Box 1) illustrate the general result that individual dynamic events in biology that are relatively slow on a macroscopic scale (having, for example, a microsecond rate constant) actually occur very rapidly on a microscopic scale (in, for example, a picosecond). The much slower macroscopic rate is due to the rarity of the microscopic event resulting from the presence of energy barriers. Events involving a few particles are limited by activation barriers if the rate constant is on the order of 10^{10} s^{-1} or less.

The general permeability of proteins to ligands suggested by the myoglobin simulation is supported by several experiments, particularly aromatic side-chain fluorescence quenching and amide hydrogen exchange data. Observations that solvent molecules can quench the fluorescence of even deeply buried groups and that amide protons completely inaccessible to solvent in the static protein structure exchange with solvent hydrogens at a relatively high rate can be explained either by assuming the protein unfolds and refolds or that solvent can penetrate rapidly into the protein interior. Molecular dynamics simulations suggest that the latter mechanism is correct, at least at temperatures much below that at which denaturation occurs. At higher temperatures transient denaturation can be important.

The primary event of photosynthesis is another biologically important phenomenon where the motions of protein atoms seem to be essential¹⁸. Photosynthetic bacteria trap solar energy in pigments bound to light-harvesting antenna proteins. The resulting excitation is transferred to a 'special pair' of bacterial chlorophylls. The excited electron is ejected from the special pair and rapidly transferred to a bacterial pheophytin molecule 17 Å away, from whence it moves on to a quinone. Although many biological processes are slow, the primary event in photosynthesis is observed to occur on the picosecond timescale¹⁹. This is because light provides the energy necessary to trigger the activated event involved in the electron transfer. Once the

charge separation has occurred, the charged species (for example, the positive special pair and negative pheophytin) are stabilized by the femtosecond relaxation of protein polar groups (Box 2).

Structure determination and refinement

No one knows how to predict the native structure of a protein from its amino-acid sequence, unless that sequence has significant homology with another protein of known structure. It is impossible, for example, to take the extended polypeptide chain, run a molecular dynamics simulation, and have it fold automatically into the correct structure. Even if there were enough supercomputer time to run a simulation for the time required to fold up in solution (for example, a 10^{-3} s simulation uses about 10^8 hours of computer time) it is not clear whether the correct structure would be obtained, as the potential functions are not perfectly accurate and the free-energy of denaturation is very small (of the order of 0.01 kcal per atom). If, however, the potential function in equation (1) is supplemented by extra information, such as a subset of the interatomic distance constraints characteristic of the folded molecule, a simulation may be able to find the correct structure in 10 picoseconds. Box 3 describes how this approach can be used with distance constraints from NMR data to determine the structure of a protein^{20,21}. Simulations have been used also to show that the internal protein motions, which alter the instantaneous interproton

2. Photosynthetic reaction centre. The first electron-transfer step in photosynthetic bacteria has extremely high efficiency, owing to two factors. The very fast rate of forward transfer (about 1 ps) prevents de-excitation from competing with electron transfer. Of equal importance is the slow rate of back transfer (1 ns to 1 μs), which is necessary because the second step of the process, transfer from the pheophytin to the quinone, is relatively slow (200 ps). Molecular dynamics has been used to show that relaxation of the reaction centre proteins stabilizes the separated electron-hole pair and thereby can contribute to the slowing down of the back reaction rate. Such a simulation is made possible by the determination of the high-resolution X-ray structure of the reaction centre from *Rhodospseudomonas viridis*²⁷. The pigments, prosthetic groups of transmembrane proteins, are organized in a roughly linear fashion across the membrane with their redox potentials ordered so as to funnel the electrons in the appropriate direction.

Although full molecular dynamics simulations of the reaction centre are very costly due to the very large size of the complex, the stochastic dynamics method described in the text can be used if the assumption is made that transfer of the electron to and from a chromophore produces only local structural changes. Such a simulation has been carried out to compare the charge-separated state with the initial state¹⁸. No large changes in structure took place during the simulation although there are numerous small adjustments in the atom positions surrounding the charged chromophores. Analysis of the chromophore energies showed that this structural relaxation, complete in femtoseconds, altered the electrostatic interactions between the protein and the chromophores so as to stabilize the special pair and pheophytin in their charged states. With such a stabilization, which corresponds to 'solvation' of the charge pair, analogous to solvation in aqueous solution, the energy matching between the charge transfer and neutral chromophore states required for rapid electron transfer is destroyed and the back transfer rate is reduced sufficiently so that it cannot compete with efficient forward transfer to the quinone. A corresponding low temperature simulation shows that the forces due to the charges are large enough for the adjustment of atomic positions and the resultant stabilization to occur at 10 K, in accord with the experimental result that the electron-transfer efficiency is nearly temperature independent.

The computed change in electrostatic energy is the result of small motions of many residues, rather than the large displacement of one or a few charged side chains. This implies that it will be difficult to confirm the simulation results experimentally either by structural studies or site-specific mutagenesis. The photosynthetic reaction centre thus demonstrates the importance of simulations for understanding functional dynamics.

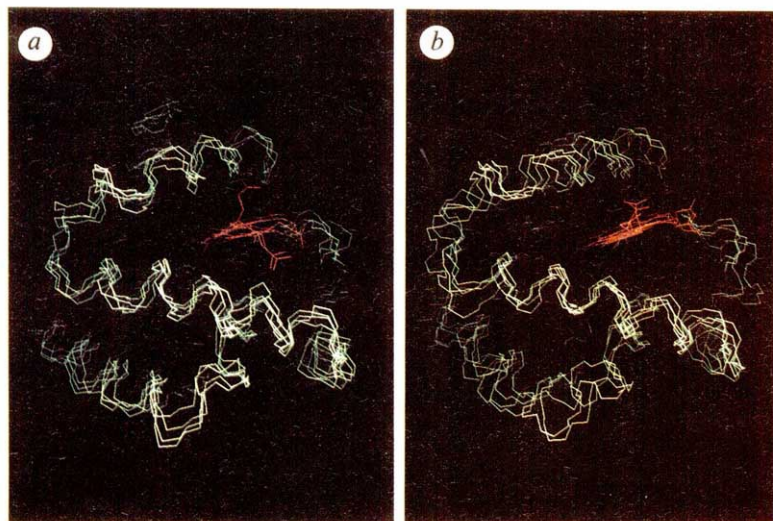


FIG. 4 *a*, Set of structures, shown as backbone traces from the myoglobin minimizations separated by about 10 ps (R. Elber and M. K., unpublished; see also ref. 13). *b*, set of structures, shown as backbone traces, from X-ray data for different globins (see refs 14, 15).

distances, usually introduce only a small error in the interpretation of the experimental data in terms of a rigid structure⁸.

In principle any information can be included as the effective potential, $E_{\text{Tot}}(\mathbf{R})$, in a molecular dynamics simulation by way of a constraint term, $E_{\text{const}}(\mathbf{R})$, such that

$$E_{\text{Tot}}(\mathbf{R}) = E(\mathbf{R}) + E_{\text{const}}(\mathbf{R}) \quad (5)$$

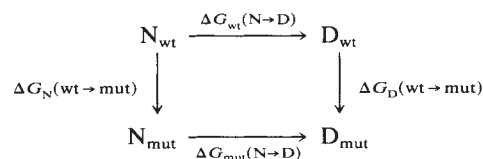
where $E(\mathbf{R})$ is the potential energy function in equation (1). One possibility might be theoretical constraints given by rules developed from the study of known protein structures. The simulations would then generate a refined structure consistent with that information. Protein crystallography, whether by X-ray or neutron diffraction, is an area where structure refinement is a major problem and where molecular dynamics with equation (5) has been shown to be able to assist in the refinement process. Box 4 outlines this important application^{22,23}.

Free-energy simulations

Molecular dynamics methods can also provide a tool for calculating free-energy differences between similar structures. Examples include the determination of the effect of a site-directed mutation on protein stability or the rate of an enzymatic reaction. A related problem concerns the effect of chemical change to a drug molecule on its affinity for the target binding site, where the relevant parameter is the free-energy difference between the bound drug and the drug in solution. Both quantities are changed when the drug is altered so that observed tighter binding could be due to a decreased solvation free energy of the modified drug rather than an increase in the free-energy of

binding to the protein target. The two types of contributions to binding and protein folding have been elucidated by free-energy simulations²⁴. Of particular importance has been the capability of free-energy simulations to determine the role of solvation and of specific protein residues to the measured free-energy changes, which generally are small and mask the individual components.

It is useful to express the free-energy change in terms of a thermodynamic cycle. For protein stability, for example, we have:



where N_{wt} , D_{wt} and N_{mut} , D_{mut} denote the wild-type and mutant proteins in the native and denatured states, respectively, and the various ΔG values in the diagram correspond to the free-energy associated with the changes of state indicated by the arrows. This type of thermodynamic cycle, an example of Hess's law, is analogous to the Born cycle used to determine the solvation energy of an ion. The measured quantity is:

$$\Delta\Delta G = \Delta G_{\text{mut}}(\text{N} \rightarrow \text{D}) - \Delta G_{\text{wt}}(\text{N} \rightarrow \text{D}) \quad (6)$$

but this corresponds to reactions that are difficult to simulate because the change from the native (folded) state to the denatured state is a complex large-scale process. To avoid this

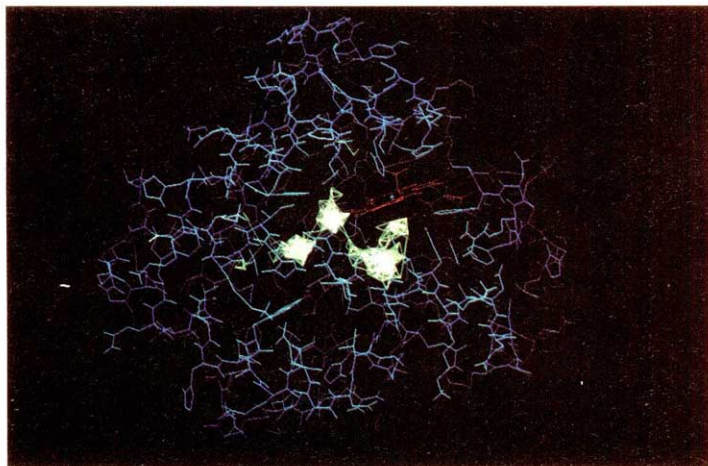


FIG. 5 Trajectory of a CO molecule escaping from myoglobin (R. Elber and M.K., unpublished).

problem, one can resort to the flexibility of the molecular dynamics methodology and do the calculation by following an unphysical path that makes use of 'computer alchemy'. Instead of changing base metal into gold (easily done in a simulation), one amino acid is changed into another. Thus, one calculates:

$$\Delta\Delta G = \Delta G_D(\text{wt} \rightarrow \text{mut}) - \Delta G_N(\text{wt} \rightarrow \text{mut}) \quad (7)$$

which, by the thermodynamic cycle, equals $\Delta\Delta G$ in equation (6).

As the desired quantities are thermodynamic functions, their evaluation by the computer 'experiment' that converts the wild type into the mutant form must be done reversibly. This is accomplished in practice by making the transformation in small increments. The free-energy change, ΔG , from state A to state B is given by:

$$\Delta G = G_B - G_A = \int_0^1 \langle \Delta V \rangle_\lambda d\lambda \approx \sum_i \langle \Delta V \rangle_{\lambda_i} \Delta\lambda_i \quad (8)$$

where

$$\Delta V = V_B - V_A \quad (9)$$

and

$$V_\lambda = (1 - \lambda) V_A + \lambda V_B \quad (10)$$

Here λ is a fractional coordinate describing the transition from the initial state ($\lambda = 0$) to the final state ($\lambda = 1$), and V_A and V_B are the empirical potential energy functions (see equation (1)) for the wild-type and mutant form, respectively.

The linear potential scaling (equation (10)) and the thermodynamic integration formula (equation (8)) make possible a decomposition of the free energy into contributions from the solvent and individual protein residues²⁵. The resulting free-energy differences include the entropic as well as the enthalpic

3. Structure determination. Two-dimensional NMR techniques are able to determine approximate short-range (less than 5 Å) interatomic distances by the nuclear Overhauser effect (NOE). To obtain an NOE the spectroscopist irradiates the protein at a frequency corresponding to the resonance of a particular proton. Resonances arising from protons within 5 Å of the target proton will have their magnitude affected by a through-space (spin dipole-spin dipole) transfer of magnetization. An NOE between two protons is therefore diagnostic of the short distance between them, even if they are far apart in the primary sequence. For a typical small protein of 50 amino acids it may be possible to obtain 300 or more approximate NOE distances to use as constraints in a folding simulation. Many of these will be between residues close together in sequence, which will help to establish the secondary structure of the protein; the others are important for determining the tertiary structure.

The NOE distance information is incorporated into the dynamics simulation as in equation (5) with $E_{\text{const}}(\mathbf{R})$ a sum of terms for the NOE distances, each of which is like the bond energy terms in $E(\mathbf{R})$. In the simplest case, a harmonic interaction is used, such that:

$$E_{\text{const}}(\mathbf{R}) = E_{\text{NOE}}(\mathbf{R}) = \sum_i A_i (R_i - R_{\text{NOE}})^2$$

Here R_i is the distance between the i th pair of protons in the dynamics structure, R_{NOE} is their distance estimated from the NOE experiment, and A_i is a 'force constant' chosen to give the forces due to the NOE terms a weight that is similar in magnitude to that of the rest of the empirical potential energy function. Thus, $E_{\text{NOE}}(\mathbf{R})$ provides a continuous energy penalty that can be differentiated to obtain a force that tends to zero as the structure satisfies the NOE distance constraints.

This method has been applied to determine solution structures of many proteins. In an early model study of the small (46 residue) protein crambin, the molecular dynamics simulations converged to the known crambin structure from several different initial extended structures (Fig. 6).

terms. They are obtained by determining the ensemble average of the potential energy differences between A and B, $\langle \Delta V \rangle_\lambda$, for the phase space corresponding to the hybrid system V_λ , and then integrating over the values of λ . By the ergodic theorem the time average behaviour obtained from a molecular dynamics simulation is equivalent to the spatial average over the configurational space in the long-time limit.

Free-energy simulations have been performed for drug-macromolecule interactions and for the effect of site-specific mutations on binding and catalysis, generally with good agreement between computationally and experimentally obtained values of $\Delta\Delta G$. The real power of the approach, however, lies in the possibility of detailed analysis of the results. Comparison of the crystal structure of a mutant protein with that of the wild type can often provide experimental clues as to the origin of the free energy changes. But the structural changes, *per se*, do not provide a direct measure of the thermodynamic quantities of interest²⁵. This is particularly true because solvation effects

4. Structure refinement. The initial model used for solving a protein crystal structure is determined by fitting to an electron density map that usually contains numerous ambiguous regions and some incorrect density. It is possible to improve this model by an iterative procedure that compares the observed X-ray diffraction data with those calculated by Fourier transformation of the assumed structural model. The refinement process is performed by minimizing a residual function, the sum of the squares of the differences between the observed ($|F_{\text{obs}}(h, k, l)|$) and calculated ($|F_{\text{calc}}(h, k, l)|$) structure factor amplitudes. Refinement progress is monitored by the value of the R -factor,

$$R = \frac{\sum_{h,k,l} ||F_{\text{obs}}(h, k, l)| - |F_{\text{calc}}(h, k, l)||}{\sum_{h,k,l} |F_{\text{obs}}(h, k, l)|}$$

where h, k, l are the reciprocal lattice points of the crystal and $||$ indicates magnitudes. Protein crystallographers have used non-linear least squares methods to minimize R . As this has a small radius of convergence, repeated cycles of time-consuming manual repositioning of the atoms that are too far from their correct positions and least-squares minimization are required.

The crystallographic residual function is strikingly similar to the NOE constraint term, suggesting that a dynamics calculation carried out with equation (5) and

$$E_{\text{const}} = E_{\text{Xray}}(\mathbf{R})$$

where

$$E_{\text{Xray}}(\mathbf{R}) = S \sum_{h,k,l} (|F_{\text{obs}}(h, k, l)| - |F_{\text{calc}}(h, k, l)|)^2$$

would produce a structure with good stereochemistry that also satisfied the observed X-ray data; the factor S is chosen to balance the two terms in equation (5). Use of molecular dynamics in a simulated annealing mode²² extends the radius of convergence of this refinement procedure and reduces the amount of manual intervention required. In simulated annealing, the protein is heated to a high temperature (up to 10,000 K) to allow it to traverse a wide range of coordinate space and overcome the local minimum problem before cooling it down to 'anneal' the structure to the lowest energy state. It is easy to apply this methodology in a molecular dynamics simulation as velocity scaling allows the simulation to be carried out at any effective temperature. After minimization of the initial model to remove any bad interatomic contacts, a dynamics simulation with the X-ray constraint term is carried out for several picoseconds at high temperature. The X-ray constraints pull the structure toward agreement with the data, while the energy function prevents it from assuming impossible geometries. The protein samples many different conformations in a few picoseconds. After the dynamic search, the protein is cooled during a further short simulation, and finally the structure is energy-minimized to ensure good stereochemistry. As in the case of the NMR analysis (Box 3), simulated annealing was first tested on crambin and has now been successfully applied to the X-ray refinement of numerous proteins²³.

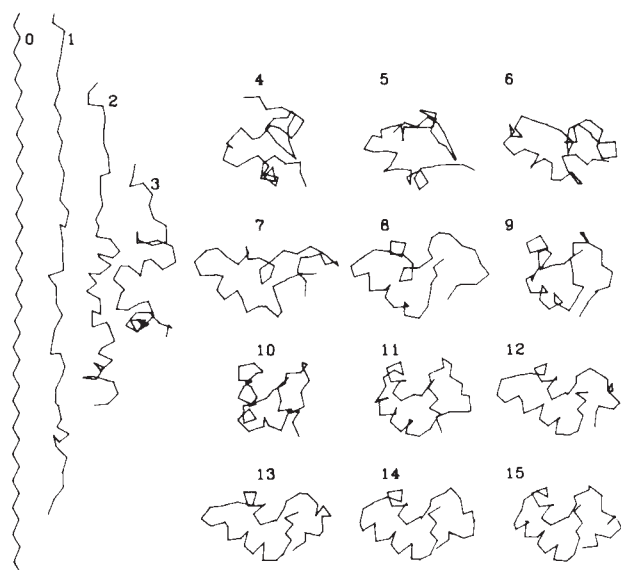


FIG. 6 Folding of crambin backbone by restrained molecular dynamics. Starting with the extended initial structure (0) snapshots are shown at 1-ps intervals; only the C α atoms are shown (from ref. 20).

are difficult to determine and entropic, as well as energetic, effects can be important. Simulations are necessary to determine which of the structural changes are important and to evaluate the entropic contributions. Insights provided by simulations are illustrated by the effect on cooperativity of the mutation of residue Asp 99 to Ala in the β subunit of haemoglobin (Box 5).

Limitations and prospects

Two limitations in existing simulations are the approximations in the potential energy functions and the lengths of the simulations. The first introduces systematic errors and the second, statistical errors. As the questions that are being asked have changed from those concerned with the qualitative characterization of the dynamics to more quantitative determinations of thermodynamics and kinetics, improved accuracy based on refined potential functions is needed. Much effort is being expended by many groups to improve the form of the potential functions and to obtain more accurate parameters. There is also a shift from using gas-phase experiments and *ab initio* calculations for isolated molecules toward parametrization with emphasis on solution and crystal data. More complex bonding terms are being used and polarization and higher-than-monopole terms are being introduced for the electrostatic interactions. Improved evaluations of the long-range electrostatic contributions are being made by the use of Ewald sums in crystals and multistep algorithms with multipole expansions for inhomogeneous systems. The combination of such refinements with the solution of the Poisson-Boltzman equation²⁶ may permit the inclusion of the dominant effect of solvation (including ionic strength) and reduce the need for full simulations with explicit solvent. But here, as in other attempts to use continuum models, the frequency dependence (of the dielectric constant, for example) must be considered in dynamic versus thermodynamic applications. For reactions, quantum calculations can be used to obtain the forces on the parts of the system undergoing bond breaking or bond making. Moreover, where necessary, quantum dynamics can be introduced, particularly for thermodynamic (high-frequency vibrations) and reaction studies (tunnelling) where classical approximations to the dynamics break down.

For many problems, simulations of 100 picoseconds are sufficient. Longer simulations would, however, be useful in the study of the relaxation phenomena involved in NMR and fluorescence depolarization, because the experimental time scale

is on the order of nanoseconds. Extended simulations would be very useful for characterizing the motions involved so as to make the interpretation of experiments more meaningful. Of equal interest are structural changes (such as the relaxation of myoglobin after photodissociation of carbon monoxide), which could be probed directly by such simulations.

There are some biologically relevant problems at the molecular level to which simulations have so far made only very limited contributions. These include the study of large-scale conformational changes (such as that from the deoxygenated to oxygenated structure in haemoglobin), details of enzymatic kinetics, very rare events such as hydrogen exchange and the 'ultimate' problem of protein folding. In all these it is likely to be necessary, as well as expedient, to replace brute force dynamics by methodologies specialized for the problem under consideration.

For enzymatic reactions, for example, which tend to be localized in character, techniques analogous to those used for simple gas-phase reactions should be adequate in most cases. They are based on activated dynamics¹, whose essential element is the determination of a transition state for the reaction and the performance of trajectories starting at the transition state to obtain satisfactory statistics for evaluating the transmission coefficient. Hydrogen exchange is of interest, not because of its intrinsic functional importance but because its wide range of timescales suggests it may be a useful probe of many aspects of protein dynamics. For conformational changes a number of simplified studies (such as hinge bending, which occurs in lysozyme, and the allosteric transition in haemoglobin) that treat the protein as composed of quasi-rigid units have had some success. For haemoglobin, an analysis which treats the helices

5. Free-energy simulations. Asp 99 β is situated at the α - β interface of the haemoglobin tetramer, a region involved in the transmission of information between subunits when haemoglobin makes the transition from the deoxygenated to oxygenated state. The mutation Asp 99 β $\rightarrow$ Ala (Hb Radcliffe) decreases the cooperativity and increases the oxygen affinity of haemoglobin. On the basis of the crystal structures of deoxygenated and oxygenated forms of haemoglobin, Perutz suggested that the essential role of Asp 99 β is to stabilize the deoxygenated form, by making hydrogen bonds to two specific residues (Tyr 42 and Asn 97) of the α subunit. The experimentally determined magnitude of the free-energy difference, $\Delta\Delta G$ (3.5 kcal per mol of interface), between the deoxygenated and oxygenated states of the mutant haemoglobin relative to the wild-type protein is consistent with the interpretation that the mutant has lost two hydrogen bonds.

Free-energy simulations²⁵ reveal that the effect of changing Asp 99 β to Ala is more complex than analysis of crystal structures had suggested. The mutation destabilizes both the deoxygenated and oxygenated forms of the haemoglobin tetramer, owing in large part to the loss of solvation free energy, but the destabilization of the deoxy tetramer (66 kcal per mol of interface) is greater than the destabilization of the oxy tetramer (60.5 kcal per mol of interface). The observed $\Delta\Delta G$ is a small difference (5.5 kcal per mol of interface) between the large individual ΔG values. In addition to the solvation effect, interactions with protein residues are also significant. Tyr 42 and Asn 97 are found to be important, but a number of other residues that have not been considered previously make larger contributions.

Although the structural change between oxygenated and deoxygenated haemoglobin may make the Asp 99 β mutation unusual in its complexity, the simulation results make clear that small measured $\Delta\Delta G$ values can arise from large 'hidden' contributions. In particular, partly cancelling changes in solvent-protein interactions and electrostatic repulsive interactions that may not be evident from a crystal structure need to be considered. Further, the results suggest that the observed free-energy changes may be small even when they contain many large contributions because the plasticity in the protein-protein and protein-solvent interactions lead to approximate cancellation of the effects for mutants that are stable.

as approximately rigid bodies attached to flexible hinges would seem to be the next step.

Finally, for protein folding it may be that the best approach to the dynamics is to study unfolding, so that at least the initial state is well defined. To achieve the required timescale it may be useful to introduce simplified models, some of which have already been used in brownian dynamics simulations and energy minimization¹. As it is likely that the rate of this slow process (as of any other) is limited by barriers to metastable intermediates, it may be useful to follow the transitions in a stepwise fashion.

Many applications of molecular dynamics require only straightforward use of the standard methodology to solve problems of biological interest and aid in the interpretation and analysis of experimental data (such as NMR and X-ray diffraction data). There are a variety of problems that can clearly be

solved but that require special methods to obtain meaningful results. Finally, there are problems, some of fundamental interest, for which molecular dynamics can be useful but for which new ideas will be required (probably beyond those suggested here) before a full solution is obtained.

It is an exciting time in the rapidly developing field of molecular dynamics simulations of biomolecules. This review will have served its purpose if it stimulates even a few biophysicists and biochemists, not to mention molecular biologists, to use molecular dynamics simulations as an aid in solving their problems. □

Martin Karplus is at the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138, USA, and Gregory Petsko is at the Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254, USA.

- Brooks, C. L., III, Karplus, M. & Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics* (Wiley, New York, 1988).
- McCammon, J. A. & Harvey, S. *Dynamics of Proteins and Nucleic Acids* (Cambridge University Press, Cambridge, 1987).
- Barnes, J. E. *Nature* **338**, 123-126 (1989).
- Brooks, B. R. *et al.* *J. comput. Chem.* **4**, 187-217 (1983).
- Stillinger, F. H. & Rahman, A. *J. chem. Phys.* **60**, 1545-1557 (1974).
- McCammon, J. A., Gelin, B. R. & Karplus, M. *Nature* **267**, 585-590 (1977).
- Aqvist, J., van Gunsteren, W. F., Leijonmark, M. & Tapia, O. *J. molec. Biol.* **183**, 461 (1985).
- Dobson, C. M. & Karplus, M. *Meth. Enzym.* **131**, 362-389 (1986).
- Gelin, B. R. & Karplus, M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 801-805 (1977).
- Ringe, D. *et al.* *Trans. Am. cryst. Ass.* **20**, 109-122 (1984).
- Doster, W., Cusack, S. & Petry, W. *Nature* **337**, 754-755 (1989).
- Smith, J., Kuczera, K. & Karplus, M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1601-1605 (1990).
- Elber, R. & Karplus, M. *Science* **235**, 318-321 (1987).
- Chothia, C. & Lesk, A. M. *Trends biochem. Sci.* **10**, 116 (1985).
- Lesk, A. M. & Chothia, C. *J. molec. Biol.* **136**, 225 (1980).
- Case, D. A. & Karplus, M. *J. molec. Biol.* **132**, 343-368 (1979).
- Elber, R. & Karplus, M. *J. Am. chem. Soc.* (in the press).
- Teutlein, H. *et al.* in *The Photosynthetic Bacterial Reaction Center* (ed. Vermeglio, A.) 101-118 (Plenum, London, 1988).
- Fleming, G. R., Martin, J.-L. & Breton, J. *Nature* **333**, 190-192 (1988).
- Brünger, A. T., Clore, G. M., Gronenborn, A. M. & Karplus, M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3801-3805 (1986).
- Wüthrich, K. *Accs chem. Res.* **22**, 36-44 (1989).
- Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458-460 (1987).
- Brünger, A. T. & Karplus, M. *Accs chem. Res.* (in the press).
- Beveridge, D. L. & Di Capua, F. M. A. *Rev. Biophys. biophys. Chem.* **181**, 431-492 (1989).
- Gao, J., Kuczera, K., Tidor, B. & Karplus, M. *Science* **244**, 1069-1072 (1989).
- Harvey, S. *Proteins* **5**, 78-92 (1989).
- Deisenhofer, J. & Michel, H. *Science* **245**, 1463-1473 (1989).
- Post, C. B. *et al.* *J. molec. Biol.* **190**, 455-479 (1986).
- Hoch, J. C., Dobson, C. M. & Karplus, M. *Biochemistry* **24**, 3831-3841 (1985).

ACKNOWLEDGEMENTS. We are most grateful to our students, postdoctoral fellows and senior colleagues, many of whom are listed in the references, for their contributions to the work described in this review.

ARTICLES

The three-dimensional seismic velocity structure of the East Pacific Rise near latitude 9° 30' N

Douglas R. Toomey^{*§}, G. M. Purdy[†], Sean C. Solomon^{*} & William S. D. Wilcock[‡]

^{*}Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[†]Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

[‡]MIT/WHOI Joint Program in Oceanography, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Three-dimensional images of crustal seismic structure beneath the East Pacific Rise show pronounced axial heterogeneity over distances of a few kilometres. A linear high-velocity anomaly, approximately 1-2 km in width and restricted to the uppermost 1 km of the crust, is centred on the rise axis. An axial low-velocity anomaly at depths of 1-3 km varies in amplitude along the axis, consistent with a zone of higher crustal temperatures midway between two discontinuities in the morphology of the rise axis. This apparent thermal segmentation along axis is consistent with injection of mantle-derived melt midway along a locally linear, 12-km-long segment of the rise.

tion of mantle-derived melt midway along a locally linear, 12-km-long segment of the rise.

OVER the past two decades our view of sea-floor spreading has been transformed from being two-dimensional and time invariant to three-dimensional and time-dependent¹⁻⁸. Most current hypotheses for the well documented along-axis variability in the morphology and geology of mid-ocean ridges invoke some form of mantle and crustal-level magmatic segmentation^{1,8-12}. The specific relation between sites of vertical injection of magma and tectonic segmentation of the rise axis, however, is not known. Nor is it understood whether the extent of lateral crustal-level magma migration controls axial discontinuities in sea-floor morphology such as overlapping spreading centres or deviations from axial linearity^{8,12,13}. Such issues may be addressed by mapping in detail the characteristics of both axial thermal structure and sea-floor morphology, enabling the correlation of morphological features with the structure of underlying magmatic systems. Because magmatism and thermal structure are

§ Present address: Department of Geological Sciences, University of Oregon, Eugene, Oregon 97403, USA.

intimately related, seismic velocity—a physical property that is strongly dependent on temperature—can provide important constraints on the structure of axial magmatic systems. Here we present tomographic images of the three-dimensional structure of the crustal seismic velocity of a sea-floor spreading centre.

The seismic imaging experiment, conducted from the R/V *Thomas Washington* during January 1988 as a cooperative project by the Woods Hole Oceanographic Institution (WHOI) and the Massachusetts Institute of Technology (MIT), was centred on the East Pacific Rise (EPR) at latitude $9^{\circ}30'N$, more than 60 km south of the Clipperton fracture zone (Fig. 1). The local half-rate of spreading is 56 mm yr^{-1} (ref. 14). Several investigators have interpreted this section of the EPR, between the Clipperton fracture zone to the north and the $9^{\circ}03'N$ overlapping spreading centre to the south, as having a relatively large axial magma chamber. A broad rectangular summit⁴, shallow average axial depths⁴, a well developed summit graben¹⁵, a bright upper-crustal reflector attributed to high concentrations of melt at the top of an axial magma chamber (AMC)^{6,16}, and high average quenching temperatures of erupted basalts⁸ have been cited as evidence of this. Multi-channel seismic reflection records have been interpreted to indicate that the AMC reflector south of the Clipperton fracture zone is continuous along the axis for $\sim 90\text{ km}$ (ref. 6), the distance separating the tectonic features that segment the rise axis.

Previous experiments have yielded one- and two-dimensional representations of the average seismic structure beneath the axis of the EPR^{17–24}, as well as information on the depth, width and along-axis continuity of the upper-crustal reflector interpreted to be the top of the AMC^{6,16,25,26}. Our image of compressional wave (P-wave) velocity in a $16 \times 16 \times 5\text{ km}^3$ volume clearly indicates that the axial crustal structure of the EPR is three-dimensional. The uppermost 1 km of the crust undergoes considerable evolution in physical properties, and the responsible processes vary along the axis. At depths of 1–3 km there is a local segmentation and asymmetry of the axial low-velocity zone that is strikingly similar in extent and dimensions to the segmentation and asymmetry of axial morphology. These tomography results provide important constraints on models for along-axis variations in mid-ocean ridge magmatism.

Experiment geometry and setting

The configuration of the tomography experiment is shown in Figs 1 and 2. Fifteen ocean-bottom instruments, eight analog

and five digital WHOI ocean-bottom hydrophones and two digital MIT ocean-bottom seismometers, recorded energy from 480 explosive shots and several airgun profiles, yielding more than 12,000 seismic records. The explosive shots (Fig. 1) were of uniform size (54.5 kg), construction and depth of detonation, ensuring a repeatable source signature. Most of the shots were in an $18 \times 16\text{ km}^2$ area centred on the rise axis; the nominal shot spacing was 1 km, except that within 3 km of the axis the shot separation was decreased to $\sim 0.5\text{ km}$. In addition to seismic data, we collected Sea Beam bathymetry, gravity and scalar magnetic field data over a $3,600\text{-km}^2$ area centred on the axis.

Tomography requires that the experimental geometry be well known. The receiver coordinates were estimated from a formal inversion of direct water-wave arrival time and navigational data²⁷. This method yields estimates of relative instrument positions accurate to within 10–20 m. The relocated receiver positions, together with travel time and navigational data were used to invert for the coordinates and origin times of all shots. Here we use data from a subset of well located shots for which the parameters are constrained by at least three water-wave arrival times (Fig. 1); 80% of the shots satisfied this criterion.

A well navigated, high-resolution bathymetric map is an important component of the experiment. We corrected the Sea Beam navigation using a formal inversion method²⁸ modified to include the acoustically located shot positions as extra constraints. Including shot positions in the inversion is advantageous because their relative positions supply more precise constraints ($\pm 10\text{ m}$) than either Global Positioning Satellite (GPS) fixes ($\pm 100\text{ m}$) or bathymetric data at ship-track cross-overs ($\pm 30\text{ m}$ east-west; $\pm 60\text{ m}$ north-south). This modification also ensured that bathymetric and seismic data were registered in the same reference frame. In the area of dense shooting the Sea Beam data and shot positions were thus co-registered to within 30-m accuracy. The Sea Beam bathymetric map resulting from re-navigation and back-projection gridding²⁹ is shown in Fig. 2.

Figure 2 shows that the experiment encompassed a 12-km-long linear rise segment trending north-south and two previously undiscovered deviations from axial linearity (devals), at latitudes $9^{\circ}28'N$ and $9^{\circ}35'N$, where the local rise trend changes abruptly by about $10\text{--}15^{\circ}$. A close inspection of Fig. 2a shows that this region of the EPR is punctuated by devals at intervals of about 10–15 km, giving the rise crest a piecewise linear appearance. Although the entire axial region shown in Fig. 2 was

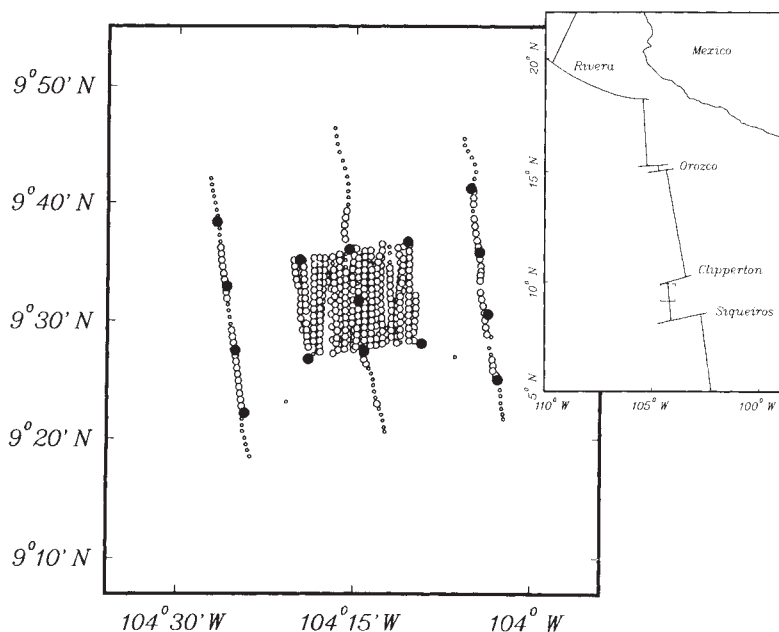


FIG. 1 Configuration of the East Pacific Rise seismic tomography experiment. The positions of the 15 ocean-bottom receivers (closed circles) and the 480 54.5-kg explosive sources (open circles) are shown; the larger open circles denote the 375 sources used here. After inversion for receiver parameters the r.m.s. residual of the water-wave arrival times is 6 ms, indicating small uncertainties in impulsive P-wave arrival times. For the 375 shots, the estimated coordinates and origin times are accurate generally to within 10 m and 4 ms, respectively, at the 50% confidence level. The experimental geometry is designed to image the crust and upper mantle in a $18 \times 16\text{ km}^2$ area centred on the rise axis. Dotted area in the inset shows the location of the experiment.

previously mapped using either Sea Beam or SeaMARC II sonar systems⁴, only the deval near 9° 20' N was previously noted⁸.

Imaging procedure

Compressional-wave arrival times are the primary data in delay-time tomography. Although the signature of the explosive source is effectively constant, the propagation medium is highly variable, giving rise to a wide variety of recorded waveforms. Many of the complex waveforms are associated with ray paths that penetrate the axial low-velocity zone and are characterized by reduced amplitudes. We developed an automatic picking routine that uses waveform alignment and a simulated annealing formalism³⁰ to enhance the signal-to-noise ratio of initial P-wave energy; the P-wave arrival time was identified using empirical functions³¹. This method can identify first-arriving P waves that sometimes precede visually identifiable arrivals by as much as 100 ms. We picked ~4,500 P-wave arrival times using this method. All arrival times were inspected, and erroneous picks were corrected.

The travel-time data were analysed for the effects of intrinsic anisotropy. In a study of the two-dimensional seismic velocity structure beneath the East Pacific Rise axis near 12° 50' N, an apparent anisotropy was estimated from the residuals at three locations of P-wave travel times relative to a one-dimensional structure²². Such residuals were consistent with an anisotropy characterized by a $\cos(2\theta)$ dependence, where θ is the azimuth relative to the trend of the rise axis. We conducted a similar analysis of the P-wave travel-time data from our fifteen receivers, but found no evidence for intrinsic anisotropy. In particular, a comparison for each receiver of maps of the travel-time residuals shows patterns that are strongly dependent on shot location irrespective of the azimuth of energy propagation.

We used the tomographic imaging technique of Thurber³² to invert for a continuously varying three-dimensional velocity structure. Velocity values are defined on a three-dimensional matrix of nodes. Between nodes, the velocity is linearly interpolated. In an $18 \times 16 \times 5$ km³ volume centred on the rise the nodal interval is 1 km in depth and across the axis, and 2 km along the axis; beyond this volume a sparser parameterization was used. In the model the crust is overlain by a uniform water layer. For the crustal path, the ray path and travel time were calculated by the pseudo-bending method³³. The crustal ray was pieced together with a water path segment from source to sea floor to yield the total travel time and propagation path. The correct water path is found by searching the Sea Beam Bathymetry for the ray entry point at the sea floor that yields a stationary minimum time path.

Our tomographic images were obtained by inverting a total of 4,278 travel-time observations from 375 shots; this set constitutes 70% of the available travel-time data from explosives. The *a priori* uncertainty in travel-time data for impulsive P-wave arrivals was assumed to be 10 ms, on the basis of the final r.m.s. residual from the inversion of water-wave arrival times. Travel times of emergent arrivals have larger uncertainties; consequently these arrivals are assigned weights of 0.75 or 0.5 that of impulsive arrivals, depending on the wave character. The weighted r.m.s. residuals of the P-wave travel times after inversion is 42 ms, significantly larger than the assumed travel-time uncertainty of 10 ms. The nodal configuration used is probably inadequate to represent shorter-wavelength crustal heterogeneity, thus giving rise to a final residual that exceeds the data uncertainty. We plan to carry out inversions using different methods of parameterization and inversion to address this issue.

Tomographic images

Map view. Several map-view sections through the three-dimensional seismic velocity image are shown in Fig. 3 at intervals of 1 km between 0 and 3 km beneath the average sea-floor

depth in the network (2,880 m). In Figs 3–5 the velocity perturbations are plotted relative to the average velocity–depth function calculated from the final three-dimensional structure.

Near the sea floor (Fig. 3a) the prominent anomaly is a narrow axis-parallel zone of high velocities. This section shows a weighted average of velocity over the uppermost kilometre of basement; for any given horizon of nodes, the weights linearly decrease with vertical distance from the horizon from 1 to 0 within 1 km (ref. 32). The ribbon of high velocities along the axis persists across the network, including the sites of devals (Fig. 2b), but the anomaly varies in magnitude and width. Seismic velocities within the anomaly are, on average, 4.6 km s^{-1} . Our results demonstrate the axial continuity and variability of this near-sea-floor anomaly. In conjunction with other seismic data^{18,20,22–24,34}, our results suggest that a laterally heterogeneous high-velocity anomaly may be a common feature of the shallow crust at the EPR axis.

Within 2–4 km of the summit the shallow axial high-velocity anomaly is flanked by an abrupt decrease in velocities of 0.6 – 1.2 km s^{-1} (Fig. 3a). In some areas this decrease occurs over distances less than 1 km. These results are consistent with an age dependence of velocities within the uppermost kilometre, consisting mainly of erupted basalts, that is characterized by a large decrease within a few tens of thousands of years after crustal formation, a characterization that holds for the entire experimental area. The image resolves an along-axis heterogeneity superimposed on this average age-dependent evolution, with differences parallel to the rise exceeding 0.4 km s^{-1} ; this degree of rise-parallel heterogeneity in the upper crust is consistent with the findings from previous off-axis refraction experiments^{35,36}.

Map-view cross-sections at 1, 2 and 3 km depth (Fig. 3b–d) clearly show a three-dimensional low-velocity anomaly beneath the rise axis. At a depth of 1 km the axial volume within ± 3 km of the rise crest is generally characterized by high average seismic velocities (6.1 km s^{-1}), but the uppermost reaches of the axial low-velocity volume (LVV) are discernible toward the south central part of the experiment area, where velocities are 0.6 to 0.8 km s^{-1} lower than in the immediately surrounding structure (Fig. 3b). The volume of relatively low velocity underlies the western portions of the high-velocity anomaly observed at the sea floor. At 1 km depth, however, the LVV is not axially continuous; instead, it terminates near morphological discontinuities (Figs 2b, 3b). The range of axial seismic velocities (~ 5.75 – 6.5 km s^{-1}) at 1 km depth in this image, together with the multi-channel seismic results interpreted to indicate an AMC reflection 1.6 km below the rise summit^{6,24}, suggest that this section depicts velocity heterogeneity in the sheeted dyke complex. Off axis the average seismic velocity and the rise-parallel velocity contrast are ~ 5.6 and 0.4 km s^{-1} , respectively. Because of topographic relief the off-axis region of the gridded model at 1 km depth depicts velocity averaged over both the sheeted dyke complex and a portion of the overlying extrusive basalts.

A section at 2 km depth beneath the datum (Fig. 3c) shows that through the seismic network the LVV is an axially continuous feature. Relative to the high-velocity anomaly imaged near the sea floor (Fig. 3a), the LVV is offset ~ 1 km to the west. The depth of this cross-section is ~ 2.3 km beneath the rise summit and thus well below the AMC reflector^{6,24}. Because our model parameterization is currently limited to nodal intervals of 1 km we are presently unable to resolve finer structure. Figure 3c shows that the magnitude of the low-velocity anomaly is modulated to produce a northern and a south central segment. The southern low-velocity anomaly, ~ 4 km wide, is a clear continuation of the relatively low velocities observed at 1 km depth. Neither the southern nor the northern low-velocity volumes extend across the morphologically defined devals at latitudes 9° 28' N and 9° 35' N (Fig. 2b). The maximum velocity contrast at 2 km depth exceeds 2 km s^{-1} ; in the along-axis direction the heterogeneity exceeds 0.5 km s^{-1} .

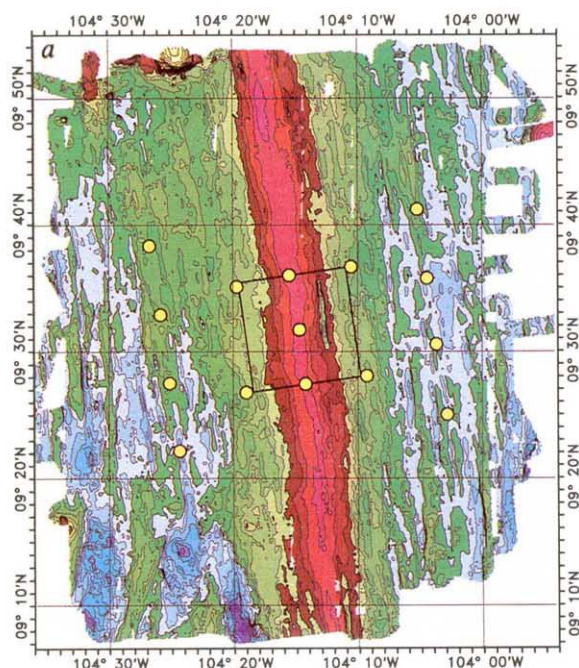
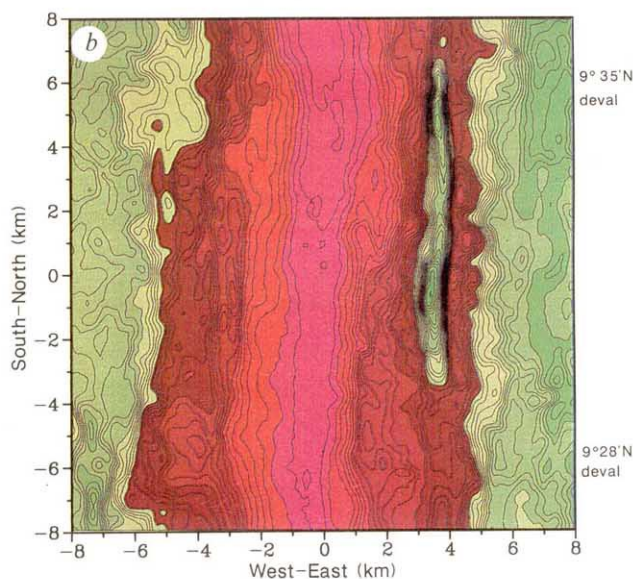


FIG. 2 *a*, A gridded bathymetric map compiled from Sea Beam data collected in a 3,600-km² area centred on the site of the tomography experiment. The positions of the 15 ocean bottom receivers (yellow disks) and the area overlying the tomographically imaged volume (box) are shown. The Sea Beam coverage is complete in the central area; renavigated Sea Beam data have been back-projected²⁹ onto a fine grid (80 × 80-m² bins). Sea floor not examined is shown in white; none of the contours have been constructed by interpolation between swaths. The contour interval is 50 m. The colour scale denotes depth as follows: yellows, <2,500 m; reds, 2,500–2,750 m; greens, 2,750–3,000 m; blues, 3,000–3,450 m. Quantitative renavigation of the ship's track (see text) precisely registers the bathymetric data. The rise summit is piecewise linear on a local scale; morphological discontinuities occur every 10–15 km. *b*, A more detailed representation of the Sea Beam bathymetry in the area of the tomographic images. The colour scale is identical to Fig. 2*a*; the contour interval is 10 m. The locations of discontinuities in the trend of axial morphology (devals) near 9° 28' N and 9° 35' N are shown. The cross-axis asymmetry of the rise summit is indicated by steeper average slopes to the east.



The section at 3 km depth (Fig. 3*d*) also shows an axially continuous low-velocity anomaly that is greatest beneath the segment of the LVV in the south central region of the experiment. At this depth the total range of perturbations in velocity also approaches 2 km s⁻¹ and the along-axis contrast is ~0.5 km s⁻¹. Similar to the sections at 1 and 2 km depth the image at 3 km depth suggests that morphological discontinuities observed at the sea floor demarcate modulations in the LVV.

The crustal velocity structure beneath 3 km depth is not well resolved in these tomographic images, which have been obtained from only a subset of the travel-time data set. Future inversions incorporating the remaining 30% of the explosives data, which includes many longer-range and thus more deeply penetrating ray paths, will permit imaging of structure at depths greater than 3 km.

Cross-sections. A comparison of closely spaced vertical cross-sections transverse and parallel to the rise axis also shows contrasting structural heterogeneity. Figure 4 shows two cross-axis sections separated by 4 km. The northern section lies 1–2 km south of the 9° 35' N deval whereas the southern section lies midway along the linear rise segment and near the crossing point of a multi-channel seismic reflection line⁶. The rapid lateral evolution of seismic structure within the upper 1–2 km of crust is apparent in both cross-sections. The lowest velocities are concentrated toward the shallowest depths in the low-velocity volume. An important observation from our experiment is that the structure in the axial low-velocity volume varies substantially along the axis. In particular, the velocities differ at 2 km depth, with the low-velocity anomaly in the southern section clearly having a larger amplitude.

A comparison of two vertical cross-sections parallel to the rise axis and 1 km to the east and west of the rise summit, as defined by the regional morphology, shows the offset of the LVV to the western side of the rise axis and the segmentation of the

axial velocity structure. The cross-section 1 km to the west of the rise summit nearly bisects the axial low-velocity anomaly and shows a marked amplification of anomalously low velocities near 2 km depth (Fig. 5*a*). In contrast, in the same depth interval a cross-section 1 km east of the rise summit (Fig. 5*b*) shows relatively higher velocities. Finally, the two axis-parallel cross-sections clearly show that the LVV is variable along axis and that the magnitude of the velocity anomaly in the LVV is at a minimum near the 9° 28' N and 9° 35' N devals.

axial velocity structure. The cross-section 1 km to the west of the rise summit nearly bisects the axial low-velocity anomaly and shows a marked amplification of anomalously low velocities near 2 km depth (Fig. 5*a*). In contrast, in the same depth interval a cross-section 1 km east of the rise summit (Fig. 5*b*) shows relatively higher velocities. Finally, the two axis-parallel cross-sections clearly show that the LVV is variable along axis and that the magnitude of the velocity anomaly in the LVV is at a minimum near the 9° 28' N and 9° 35' N devals.

Discussion

The observed segmentation of the low-velocity anomaly at 1–3 km depth (Figs 2*b*, 3*b–d*) has important implications for the thermal and magmatic structure of the rise. If we assume that the axial LVV is principally the result of elevated temperatures then the tomographic images map a segmented thermal structure. In this view, the temperature at mid-crustal depths is highest in the centre of the morphologically defined 12-km-long linear rise segment and lowest at the segment ends. The bulk seismic velocities imaged in the LVV average 5–6 km s⁻¹, or 1–2 km s⁻¹ lower than normal. Laboratory measurements³⁷ of compressional velocities in igneous rocks are consistent with a decrease of 1–2 km s⁻¹ as the temperature increases from room temperature to the solidus; much of the crustal LVV may thus be at subsolidus temperatures. Because the vertical nodal spacing in the images is 1 km we cannot rule out the presence of a thin sill of magma; such a body, however, would be only a small fraction by volume of the overall crustal low-velocity anomaly. The best evidence for the presence of a crustal body of magma at this site comes from the multi-channel seismic studies^{6,24} that indicate the presence of liquid about 1.6–2 km beneath the rise summit with a bulk compressional velocity of ~3 km s⁻¹. The liquid is believed to be confined to a thin (<1 km) sill^{24,26}, but the thickness is not well resolved. Tomographic images of the

along-axis variations in velocity near the presumed magma sill at 2 km depth (Fig. 3c) are consistent with along-axis temperature variations of perhaps 50–200 °C (ref. 37). This temperature range is subject to large uncertainties, and along-axis variations in the fraction of partial melt or pooled magma volume cannot be quantified. The inference that a higher melt concentration may be present at an elevated temperature is, however, reasonable. The tomographic images, in conjunction with the multi-channel seismic data, are thus consistent with a systematic axial variation in temperature and perhaps in the volume of magma in a thin sill embedded near the top of a larger LVV.

A segmented thermal structure may result from axial variations either in hydrothermal cooling or in the kinematics of

magma injection. Morphological segmentation of the rise, in contrast, is likely a result of either tectonic or magmatic processes. The similarities in the extent and dimensions of thermal and morphological segmentation (Fig. 2b) collectively argue that magmatic, rather than hydrothermal, processes are responsible for the observed patterns. The variation in the shallow temperature structure in the rise segment is most simply explained as the result of vertical injection of mantle-derived melt into the crust near the segment centre followed by lateral migration of melt at crustal levels toward the segment ends. Our results support the hypothesis that discontinuities in rise-axis morphology arise from the misalignment of pulses of magma that migrate along axis^{4,8,12,13}. By this interpretation, the vertical and horizontal transport of melt controls, to a large degree, axial

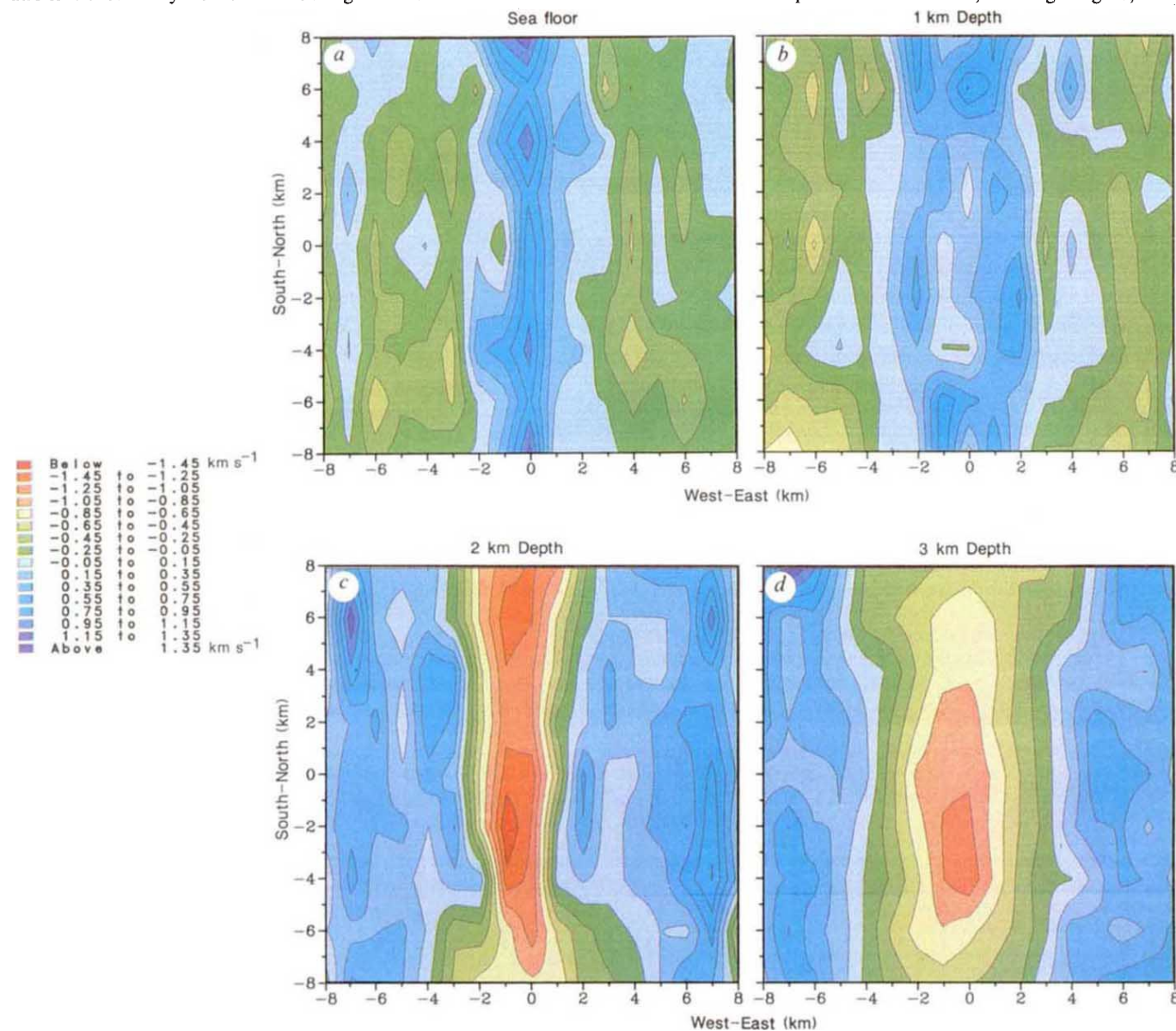


FIG. 3 Horizontal cross-sections through a three-dimensional P-wave velocity structure resulting from tomographic inversion. *a–d* Sections at 1-km intervals between 0 and 3 km depth beneath a plane 2,880 m below sea level. The velocity model is represented by perturbations from an average one-dimensional model over the area. At intervals of 1 km between 0 and 5 km depth this one-dimensional reference model is defined by the following values of velocity: 4.0, 5.8, 6.5, 6.5, 6.7 and 7.0 km s⁻¹. Between nodal locations velocity is linearly interpolated in the three spatial coordinates³². The plan-view origin for all of these plots coincides with the centre of the shooting grid (9° 31.72' N, 104° 14.66' W); the y axis is aligned with the regional trend of the East Pacific Rise (N8°W). For reference, a multi-channel reflection line⁶ at 9° 30' N crossed the rise axis ~3 km south of the centre

of the model. Along this line a wide and bright reflector in the upper crust was interpreted as the top of an axial magma chamber. The Sea Beam bathymetry in the area of the map-view sections is displayed in Fig. 2b. Before tomographic inversion the travel-time variance for a one-dimensional starting model that approximates the local off-axis crustal structure²⁴ is 0.025 s². The inversion, which is step-length dampened, terminated after nine iterations, at which point the travel-time variance was 0.002 s², yielding a variance reduction of 92%. Synthetic inversions using a nearly identical experimental geometry and a comparable three-dimensional axial anomaly demonstrate that convergence is obtained for a one-dimensional starting model and that the axial variations in the low-velocity anomaly are recoverable⁴⁹.

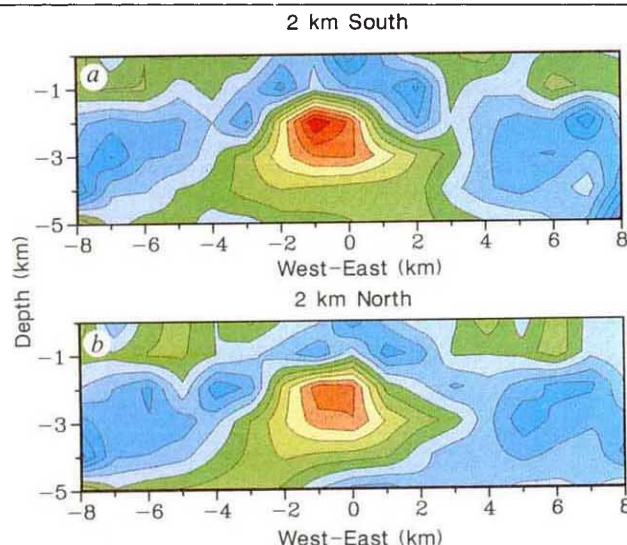


FIG. 4 Two vertical cross-sections of the P-wave velocity structure transverse to the rise axis. *a*, A section 2 km to the south of the array centre or about midway along a locally linear rise segment. This section lies near the crossing point of a seismic reflection line⁶. *b*, A section 2 km to the north of the array centre or near the distal end of a locally linear rise segment. The origin is identical to that of Fig. 3.

morphology (Fig. 2*a, b*), axial thermal structure and axial crustal seismic structure (Fig. 3*b-d*).

The segmentation of the velocity structure near 1 km depth beneath the rise summit suggests strong lateral gradients in temperature in the sheeted dyke complex, which may control the spatial concentration of axial hydrothermal activity. On axis, the sheeted dykes separate the magma reservoir from the solidified and relatively cold extrusive basalts. The seismic velocity distribution near 1 km depth (Fig. 3*b*) indicates bulk velocities as low as 5.75 km s^{-1} , below the measured velocities of unaltered dolerites at room temperatures³⁸, consistent with anomalously high shallow-crustal temperatures near the axis. If the permeability of the extrusive layer allows penetration of sea water to the depth of the sheeted dyke complex, the inferred lateral thermal gradients would favour a focusing of hydrothermal venting near the regions of lowest seismic velocities. Alternatively, the low-velocity anomaly may result from lateral variations in bulk porosity or in the degree of hydrothermal alteration³⁹.

The observation of an axial high-velocity anomaly within 1 km of the sea floor flanked by an abrupt cross-axis decrease in velocities of $0.6\text{--}1.2 \text{ km s}^{-1}$ (Fig. 3*a*) may result from either an evolution in the physical properties of the shallow crust or variations in the thickness of geological units. A commonly held view of the upper-crustal velocity structure is that it evolves in response to progressive changes in the nature and magnitude of the porosity in the extrusive basaltic layer^{20,22,23,39–43}. Consequently, the high axial velocities are inferred to indicate relatively low bulk porosity in recently erupted basaltic pillow and sheet flows and the adjacent lower velocities are inferred to be the result of increased porosity owing to faulting^{20,22–24}. The tomographic image (Fig. 3*a*), however, indicates that seismic velocities evolve most rapidly within about a kilometre of the crest of the rise summit whereas most faulting occurs near the flanks of the axial bathymetric high⁴⁴, or 2–4 km off axis at this site⁴. In this interpretation of the evolution in upper-crustal velocities, it is implicitly assumed that layering of the crust on and off the axis is identical.

In contrast, several models of shallow crustal magmatism^{45,46} suggest that directly beneath the neovolcanic zone the sheeted dyke complex rises to shallow depths beneath the sea floor. Our favoured explanation of the narrow axial high-velocity anomaly is a relatively shallow depth of the higher-velocity sheeted dyke

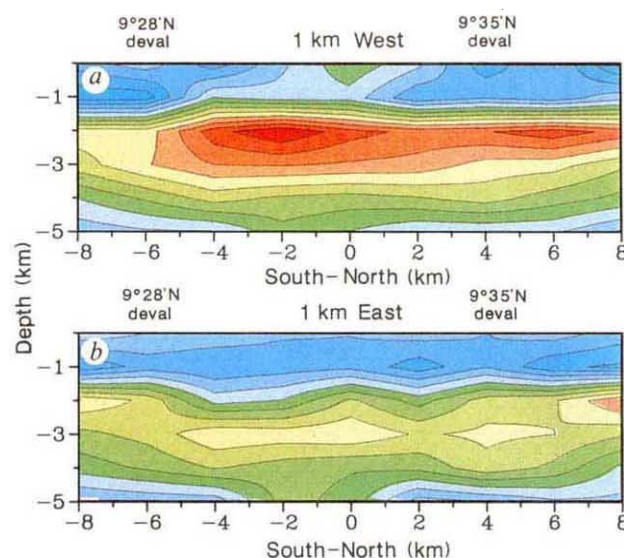


FIG. 5 Two vertical cross-sections of the P-wave velocity structure parallel to the rise axis. *a*, A section 1 km to the west of the rise crest. This section intersects the planes of section in Fig. 4*a, b* at -1 km . *b*, A section 1 km to the east of the rise summit. This section intersects the planes of section in Fig. 4*a, b* at $+1 \text{ km}$. The locations of the $9^{\circ}28' \text{ N}$ and the $9^{\circ}35' \text{ N}$ devals are shown.

complex at the locus of eruption and a rapid increase in the thickness of lower-velocity extrusives away from the neovolcanic zone. In this hypothesis individual volcanic flow units erupting from the 100–500-m-wide summit graben^{1,12} blanket a much wider area of the rise summit⁴⁷ and thus give rise to a progressive thickening of the basalt layer and a concomitant decrease in average shallow-crustal seismic velocities with age in the near-axis region.

The tomographic images and the cross-axis rise morphology indicate that axial magmatic processes are not symmetric with respect to the rise summit. The LVV between 1–3 km depth is offset about 1 km to the west from both the shallow axial high-velocity anomaly and the bathymetric rise crest (Figs 2, 3). Immediately to the south of our experiment site, several seismic reflection lines indicate a similar but more pronounced westerly offset of the AMC reflector^{6,16,25,48}. Near $9^{\circ}30' \text{ N}$ the slope of the sea floor away from the rise crest is also markedly asymmetric, with the western slope significantly less than the eastern slope (Fig. 2). This relationship, in general, holds for most of the rise between $9^{\circ}30' \text{ N}$ and $9^{\circ}10' \text{ N}$. A possible explanation of these observations is that the westward offset of crustal magmatism may be buoyantly uplifting the western rise flank.

Our preferred interpretation of the tomography results is that mantle-derived melt is injected into the shallow crust at characteristic intervals of $\sim 10 \text{ km}$ along the rise axis, giving rise to magmatically defined rise segments of similar lengths. According to this view, the scale of segmentation of the rise by crustal magma reservoirs is significantly less than that defined by either tectonic offsets of the rise or local maxima in the along-axis depth profile. Such magmatic segmentation of the rise, accompanied by lateral migration of melt at crustal levels, may explain variations in sea-floor basalt chemistry as observed from closely spaced dredge hauls⁸, along-axis discontinuities in rise morphology, and the segmentation of crustal seismic velocity structure documented here. □

Received 12 July; accepted 31 August 1990.

1. Francheteau, J. & Ballard, R. D. *Earth planet. Sci. Lett.* **64**, 93–116 (1983).
2. Schouten, H. & White, R. S. *Geology* **8**, 175–179 (1980).
3. Schouten, H. & Klitgord, K. D. *Earth planet. Sci. Lett.* **59**, 255–266 (1982).
4. Macdonald, K., Sempere, J.-C. & Fox, P. J. *J. geophys. Res.* **89**, 6049–6069 (1984).
5. Purdy, G. M. & Detrick, R. S. *J. geophys. Res.* **91**, 3739–3762 (1986).
6. Detrick, R. S. *et al. Nature* **326**, 35–41 (1987).
7. Thompson, B., Bryan, B., Ballard, R., Hanuro, K. & Melson, W. *Nature* **310**, 429–433 (1985).

8. Langmuir, C. H., Bender, J. F. & Batiza, R. *Nature* **322**, 422–429 (1986).
9. Whitehead, J. A., Dick, H. J. B. & Schouten, H. *Nature* **312**, 146–148 (1984).
10. Schouten, H., Klitgord, K. D. & Whitehead, J. A. *Nature* **317**, 225–229 (1985).
11. Crane, K. *Earth planet. Sci. Lett.* **72**, 405–414 (1985).
12. Macdonald, K. C. *et al.* *Nature* **335**, 217–225 (1988).
13. Batiza, R. & Margolis, S. H. *Nature* **320**, 439–441 (1986).
14. Klitgord, K. D. & Manmerickx, J. *J. geophys. Res.* **87**, 6725–6750 (1982).
15. Macdonald, K. C. & Fox, P. J. *Earth planet. Sci. Lett.* **88**, 119–131 (1988).
16. Herron, T. J., Ludwig, T. J., Stoffa, P. L., Kan, T. K. & Buhl, P. *J. geophys. Res.* **83**, 798–804 (1978).
17. Orcutt, J. A., Dennett, B., Dorman, L. & Prothero, W. *Nature* **256**, 475–476 (1975).
18. Rosendahl, B. R. *et al.* *J. geophys. Res.* **81**, 5244–5304 (1976).
19. Lewis, B. T. R. & Garmany, J. D. *J. geophys. Res.* **87**, 8417–8425 (1982).
20. McClain, J. S., Orcutt, J. A. & Burnett, M. J. *geophys. Res.* **90**, 8627–8639 (1985).
21. Harding, A. J. *et al.* *J. geophys. Res.* **94**, 12163–12196 (1989).
22. Burnett, M. S., Caress, D. W. & Orcutt, J. A. *Nature* **339**, 206–208 (1989).
23. Caress, D. W., Burnett, M. S. & Orcutt, J. A. *J. geophys. Res.* (in the press).
24. Vera, E. E. *et al.* *J. geophys. Res.* (in the press).
25. Mutter, J. C. *et al.* *Nature* **336**, 156–158 (1988).
26. Kent, G. M., Harding, A. J. & Orcutt, J. A. *Nature* **344**, 650–653 (1990).
27. Creager, K. C. & Dorman, L. M. *J. geophys. Res.* **87**, 8379–8388 (1982).
28. Nishimura, C. E. & Forsyth, D. W. *Mar. geophys. Res.* **9**, 333–352 (1988).
29. Stewart, W. K. *Proc. 7th Int. Conf. Offshore Mechanics Arctic Eng.* Vol. 6 (eds Chung, J. S. & Angelides, D. C.) 61–71 (Am. Soc. mech. Eng., New York, 1988).
30. Rothman, D. H. *Geophysics* **51**, 332–346 (1986).
31. Allen, R. *Bull. seism. Soc. Am.* **72**, S225–S242 (1982).
32. Thurber, C. H. *J. geophys. Res.* **88**, 8226–8236 (1983).
33. Um, J. & Thurber, C. *Bull. seism. Soc. Am.* **77**, 972–986 (1987).
34. Herron, T. J. *Geophys. Res. Lett.* **9**, 17–20 (1982).
35. Purdy, G. M. *J. geophys. Res.* **87**, 8403–8416 (1982).
36. Bratt, S. R. & Purdy, G. M. *J. geophys. Res.* **89**, 6111–6125 (1984).
37. Murase, T. & McBirney, A. R. *Geol. Soc. Am. Bull.* **84**, 3563–3592 (1973).
38. Salisbury, M. H. & Christensen, N. I. *J. geophys. Res.* **83**, 805–817 (1978).
39. Fox, P. J. & Stroup, J. B. in *The Sea Vol. 7* (ed. Emiliani, C.) 119–218 (Wiley, New York, 1981).
40. Hyndman, R. D. & Drury, M. J. *J. geophys. Res.* **81**, 4042–4052 (1976).
41. Schreiber, E. & Fox, P. J. *J. geophys. Res.* **81**, 4071–4076 (1976).
42. Houtz, R. & Ewing, J. *J. geophys. Res.* **81**, 2490–2498 (1976).
43. Purdy, G. M. *J. geophys. Res.* **92**, 9351–9362 (1987).
44. Macdonald, K. C. *A. Rev. Earth planet. Sci.* **10**, 155–190 (1982).
45. Cann, J. R. *Geophys. J. R. astr. Soc.* **39**, 169–187 (1974).
46. Kidd, R. G. W. *Geophys. J. R. astr. Soc.* **50**, 149–183 (1977).
47. ARGO-RISE Group. *Can. Mineral.* **26**, 467–486 (1988).
48. Hale, L. D., Morton, C. J. & Sleep, N. H. *J. geophys. Res.* **87**, 7707–7718 (1982).
49. Toomey, D. R. thesis, Mass. Inst. Tech./Woods Hole Ocean. Inst. (1987).

ACKNOWLEDGEMENTS. We thank the officers and crew of the R/V *Thomas Washington* and members of the scientific party for their assistance. We thank C. Nishimura for his Sea Beam renavigation software, D. Blackman, L. Menke and C. Nishimura for programming advice on Sea Beam processing, J. Foley for his contribution to the waveform alignment algorithms, J. Fox and K. Macdonald for providing copies of their bathymetric maps, and E. Vera and D. Caress for preprints. This research was supported by the NSF and the Office of Naval Research. Woods Hole Oceanographic Institution contribution No. 7463.

Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators

Isabelle Issemann & Stephen Green*

Imperial Chemical Industries PLC, Central Toxicology Laboratory, Alderley Park, Macclesfield SK10 4TJ, UK

We have cloned a member of the steroid hormone receptor superfamily of ligand-activated transcription factors. The receptor homologue is activated by a diverse class of rodent hepatocarcinogens that causes proliferation of peroxisomes. Identification of a peroxisome proliferator-activated receptor should help elucidate the mechanism of the hypolipidaemic effect of these hepatocarcinogens and aid evaluation of their potential carcinogenic risk to man.

PEROXISOME proliferators are a diverse group of chemicals that include hypolipidaemic drugs, herbicides and industrial plasticisers (see refs 1–4 for reviews). Administration of these chemicals to rodents results in the dramatic proliferation of hepatic peroxisomes as well as liver hyperplasia^{5–7}. Coincidentally, there is increased transcription of genes required for the peroxisomal β -oxidation of long chain fatty acids^{8,9} as well as genes of the cytochrome P450 IV family^{10–12}.

The main effect of the hypolipidaemic peroxisome proliferator clofibrate¹³ is to lower triglycerides and cholesterol concentrations in the plasma (see refs 14, 15 for reviews). Although beneficial in the prevention of ischaemic heart disease in people with elevated levels of cholesterol^{14,15} the pharmaceutical use of such drugs has been questioned by the demonstration that they are rodent hepatocarcinogens^{16,17}. Peroxisome proliferators are termed non-genotoxic carcinogens as they fail to cause DNA damage directly¹⁸. The increase in peroxisomal fatty acid β -oxidation seen in response to peroxisome proliferators results in a greater production of hydrogen peroxide. It has been proposed that this results in oxidative stress leading to DNA damage and possibly tumour initiation^{1,19,20}. Alternatively, or

in addition, peroxisome proliferators may act as liver-tumour promoters^{21,22}. Peroxisome proliferators therefore represent a model system to study the mechanism(s) of non-genotoxic hepatocarcinogens.

The report of a peroxisome proliferator binding protein in rat liver^{23,24}, together with the known ability of peroxisome proliferators to modulate specific gene transcription rapidly^{8,10}, suggested to us that peroxisome proliferators could act by a mechanism similar to that of steroid hormones^{25–27}. We report here the cloning of a novel member of the steroid hormone receptor superfamily activated by peroxisome proliferators.

Cloning of receptors

Comparison of the amino-acid sequence of several nuclear receptors indicated a highly conserved region within the DNA-binding domain and a mixture of 32 different 26-mer oligonucleotides was synthesized based upon the consensus sequence (Fig. 1). Approximately 700,000 phage plaques from a mouse liver complementary DNA library were screened with the oligonucleotide probe and 64 clones that consistently hybridized were grouped into 26 classes on the basis of cross-hybridization and restriction mapping data. Their sequence indicates that clone 9 is the murine thyroid hormone receptor- β (TR β) and that clones 54 (Fig. 2), 71 and 74 (data not shown) are newly discovered members of the steroid hormone receptor superfamily.

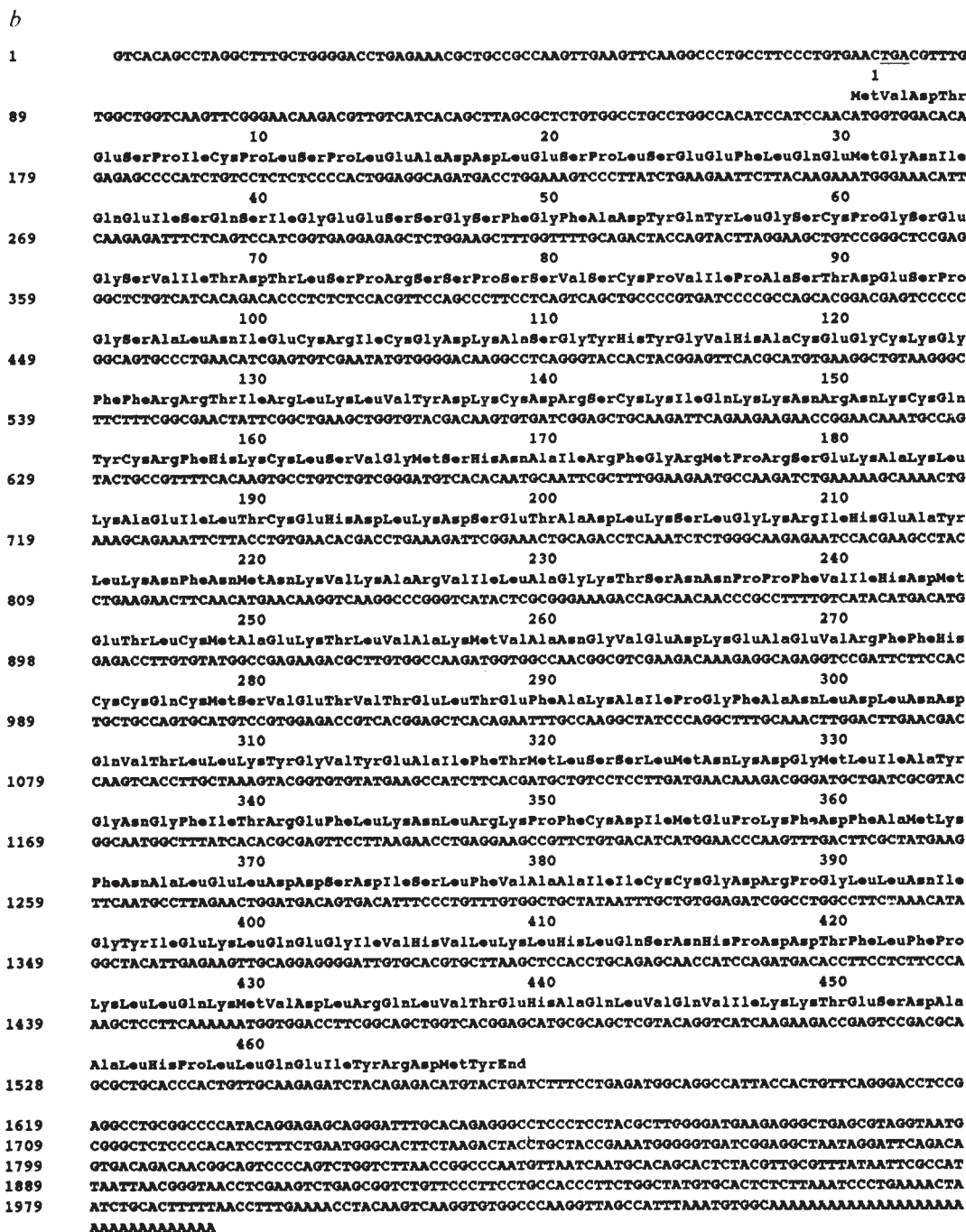
A second mouse liver cDNA library was screened using clone 54 and a new clone (clone 3) isolated that extends further 5' than clone 54 (Fig. 2a). Nucleotide sequence analysis of clones 3 and 54 indicates an open reading frame encoding a 468-amino-acid protein of relative molecular mass 52,400 (Fig. 2b; hereafter termed the mouse peroxisome proliferator activated receptor, or PPAR). The sequence surrounding the first ATG codon conforms closely to the Kozak consensus sequence²⁸ and is preceded by an in-frame translation termination codon suggesting it to be the initiator methionine codon.

Comparison of the predicted amino-acid sequence of PPAR

* To whom correspondence should be addressed.

vitamin D3 receptor (hVD3R; Fig. 3). Comparison of the putative DNA-binding domain of PPAR (amino acids 102–166) with that of hear1, E75, hRAR α , hTR β , mH2RIIBP and hVD3R indicates 64, 64, 63, 59, 58 and 46% amino-acid sequence identity, respectively. The putative ligand-binding domain (amino acids 281–468) of PPAR shares 38% identity with hear1, 33% with E75, 29% with the other receptor sequences compared here (Fig. 3) and only 21% with the human oestrogen (hER) and glucocorticoid (hGR) receptors.

To determine the putative ligand for PPAR, we examined the tissue-specific expression of the *PPAR* gene in the adult mouse. A high level of expression was seen in liver, kidney and heart with very weak expression in brain and testis (Fig. 4). Notably, two major mRNA species of about 1.8 and 2.0 kilobases (kb) were detectable in liver, whereas only the smaller mRNA was seen in kidney and heart. PPAR is also expressed in brown



METHODS. The plasmid (pBluescript SK-) inserts of the lambda zap vector were excised *in vivo* (Stratagene) and nested deletions prepared using exonuclease III (Stratagene). Double-stranded DNA was sequenced on both strands using the dideoxy-sequencing technique using T7 DNA polymerase (Sequenase; US Biochemical) and analysed using Dnastar's lasergene computer program.

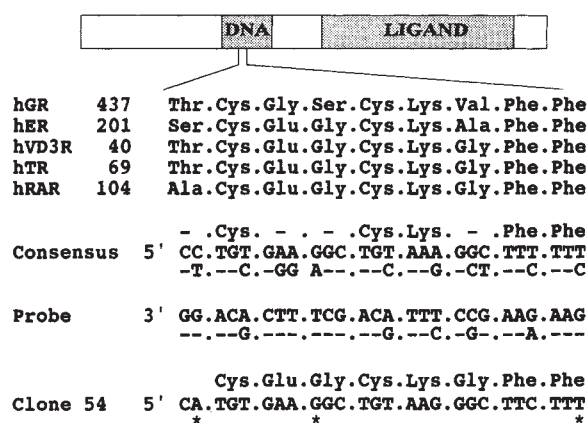


FIG. 1 Nucleotide sequence of oligonucleotide-receptor probe. The schematic structure of a steroid hormone receptor is shown at the top; DNA and ligand binding domains are shaded. Below, amino-acid sequence alignment within the DNA-binding domain of the human glucocorticoid (hGR)⁴⁰, oestrogen (hER)⁴¹, vitamin D3 (hVDR)⁴², thyroid hormone (hTR)⁴³ and retinoic-acid (hRAR)^{44,45} receptors. The consensus nucleotide and amino-acid sequence for this region and the sequence of the derived degenerate oligonucleotide probe mixture are also shown. Note that the antisense strand of the consensus sequence was used to design the probe so as to maximize G:T base-pairing in the event of a mismatch⁴⁶. The nucleotide and amino-acid sequence of clone 54 is given; asterisks, nucleotide mismatches with the probe.

METHODS. Oligo-dT-primed cDNA was synthesized using mouse (Alpk:APFCD-1) liver polyadenylated RNA and cloned into the *EcoRI*-site of lambda Zap or ZapII (Stratagene). After amplification the library was plated, transferred to duplicate Hybond N nylon filters (Amersham) and the DNA fixed with ultraviolet light. The filters were screened with the ³²P end-labelled consensus oligonucleotide-probe mixture (0.5 pmol ml⁻¹; 5 × 10⁶ c.p.m. pmol⁻¹) using low stringency conditions of hybridization—42 °C for 24 h with 0 or 10 % formamide, 1 × Denhardt's solution, 6 × NET (1 × NET is 0.15 M NaCl, 1 mM EDTA, 15 mM Tris-Cl, pH 7.5), 0.2% SDS and 100 µg ml⁻¹ denatured salmon-sperm DNA. Filters were washed 4 × 20 min at 25 °C in 2 × SSC buffer, 0.1 % SDS and autoradiographed for 3 days at -70 °C with intensifying screens.

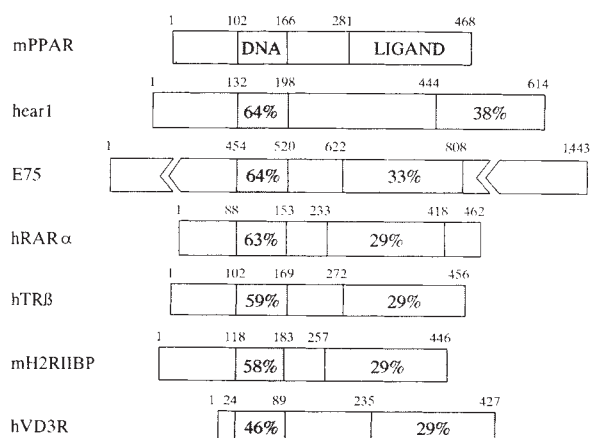


FIG. 3 Comparison of the amino-acid sequence of mouse PPAR with its nearest relatives of the nuclear receptor superfamily. The structure of each receptor is drawn schematically with the DNA and ligand binding domains shaded. Alignments were performed using Dnastar's lasergene amino-acid align program and any sequence similarity between mouse PPAR and the other receptors shown as the percentage amino-acid identity. hear1, human *erbA*-related 1 (ref 47); E75, ecdysone-inducible gene located at the E75 locus in *Drosophila*^{48,49}; hRARα, human retinoic acid receptor-α^{44,45}; hTRβ, human thyroid hormone receptor-β⁴³; mH2RIIBP, binding protein recognizing the RII region of the mouse major histocompatibility gene I (ref. 50); hVDR3R, human vitamin D3 receptor⁴².

adipose tissue (S. Plummer, P. Senior, S.G. and F. Beck, unpublished results). This pattern of expression was intriguing as it mirrors the tissue-specific effects of peroxisome proliferators^{29,30}.

Receptor activation

To test the ability of peroxisome proliferators to activate PPAR we constructed chimaeric receptor expression vectors containing regions encoding the putative ligand-binding domain of the mouse PPAR and the N-terminal sequence and DNA-binding domain of either hER or hGR (Fig. 5A; ER-PPAR and GR-PPAR). A transient *transactivation* assay was established in COS1 cells using the ER-PPAR chimaeric receptor expression vector and Vit-G-CAT reporter plasmid that contains the bacterial chloramphenicol acetyl transferase (CAT) gene under the transcriptional control of the oestrogen receptor (see Fig. 5 legend). The assay was used to screen several potential ligands. Notably, a 100 µM concentration of the hypolipidaemic peroxisome proliferator nafenopin stimulates CAT activity approximately 10–20-fold (Fig. 5b, compare lanes 6, 7). Importantly, nafenopin had no effect on CAT activity when using either the parent expression vector pSG5 (lane 2) or the hER expression vector HEO (lane 4). Similar results were observed when using the GR-PPAR expression vector together with the glucocorticoid-responsive reporter plasmid MMTV-CAT (lanes 9–16). In this system, 100 µM nafenopin induces CAT activity approximately 40-fold (compare lanes 14, 15). Thus the putative ligand-binding domain of PPAR confers nafenopin inducibility to either the oestrogen or glucocorticoid receptors in the appropriate chimaeric receptor.

We examined next the dose response of ER-PPAR to nafenopin and other peroxisome proliferators (Fig. 6a). Using CAT activity as an indicator, an ED₅₀ of about 10 µM was observed with nafenopin (Fig. 6b). Other hypolipidaemic peroxisome proliferators, Wy-14,643, methylclofenapate and clofibrate acid (the active metabolite of clofibrate¹³), had approximate ED₅₀s of 1.5, 55 and 310 µM, respectively (Fig. 6b). Mono-ethylhexyl phthalate (MEHP), a more active metabolite of the industrial plasticizer DEHP (ref. 31), also stimulated CAT activity with

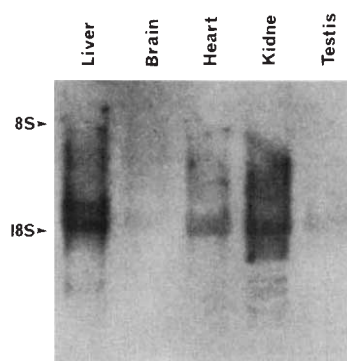


FIG. 4 Tissue-specific expression of PPAR in the mouse.

METHODS. Total RNA was prepared from adult tissues by extraction with guanidine isothiocyanate and centrifugation through a CsCl cushion⁵¹. Polyadenylated RNA was purified by oligo-dT cellulose chromatography and 10 µg fractionated on a 1.2% agarose gel containing formaldehyde⁵¹. The RNA was transferred to a Hybond N nylon filter (Amersham) and fixed with ultraviolet light. The filter was hybridized using high stringency conditions (42 °C, 50% formamide, 1 × Denhardt's, 5 × SSPE, 0.1% SDS, 100 µg ml⁻¹ denatured salmon-sperm DNA) using the ³²P random-prime-labelled (Amersham) cDNA insert of clone 54 (1–2 × 10⁹ c.p.m. µg⁻¹; 2–4 × 10⁶ c.p.m. ml⁻¹). The filter was washed briefly at 25 °C in 2 × SSPE, 0.1% SDS, 0.03% sodium pyrophosphate followed by two 20-min washes at 68 °C. Two further 20-min washes were performed at 68 °C in 0.1 × SSPE, 0.1% SDS, 0.03% sodium pyrophosphate. Autoradiography was for 60 h at -70 °C with an intensifying screen. The positions of 28S and 18S ribosomal RNA are indicated.

FIG. 5 Stimulation of CAT reporter genes by ER-PPAR and GR-PPAR chimaeric receptors in the presence of nafenopin. *a*, Schematic structure of the ER-PPAR and GR-PPAR chimaeric receptors with numbers referring to the amino-acid sequence. *b*, The oestrogen-responsive (Vit-G-CAT, lanes 1–8) or glucocorticoid-responsive (MMTV-CAT⁵², lanes 9–16) reporter genes were introduced into COS1 cells together with either the parent expression vector pSG5 (ref. 53; Stratagene) the hER expression vector HEO (ref. 41), the hGR expression vector HG1 (ref. 54) or the chimaeric receptors ER-PPAR and GR-PPAR. Cells were maintained either in DMSO vehicle (0.1% v/v) alone (–), 10^{-8} M oestradiol (E), 10^{-7} M dexamethasone (D) or 100 μ M nafenopin (N). The position of ¹⁴C-chloramphenicol (C) and acetylated ¹⁴C-chloramphenicol (AC) after thin-layer chromatography is indicated.

METHODS. ER-PPAR and GR-PPAR were constructed essentially as described previously^{52,55}. Briefly, site-directed mutagenesis was used to insert an *Xho*I restriction enzyme site into clone 54 between codons 166 and 167. To generate ER-PPAR, the *Xho*I–*Eco*RI fragment, encoding the putative ligand-binding domain of mouse PPAR, was ligated to the *Eco*RI–*Xho*I fragment of HE28, that encodes the ER DNA binding domain⁵⁵, and inserted into the *Eco*RI site of the expression vector pSG5. The fragment containing the GR DNA binding domain was constructed by insertion of an *Xho*I site into HG1 between codons 486 and 487 using 100 ng plasmid DNA and 5 cycles of PCR⁵⁶. GR-PPAR was constructed by ligating this fragment to the *Xho*I–*Eco*RI fragment of clone 54 and insertion into pSG5. Vit-G-CAT was constructed by replacing the thymidine kinase (tk) promoter of Vit-tk-CAT with that of the rabbit β -globin promoter from 17M2-G-CAT⁵². COS1 cells (3×10^5) were seeded into 9-cm-diameter plates in phenol red-free medium (Gibco) supplemented with 5% fetal calf serum (Gibco) stripped of endogenous steroids by treatment with dextran-coated charcoal. After 3 days the medium was replaced and CsCl purified plasmid DNA (10 μ g) added by calcium phosphate precipitation. Each plate contained a CAT reporter gene (1 μ g), an expression vector (pSG5, HEO, HG1, ER-PPAR, GR-PPAR; 1 μ g), the β -galactosidase expression vector, pCH110 (3 μ g; Pharmacia) and pBluescribe M13+DNA (Stratagene) as a carrier (5 μ g). Ligands (1,000 $\times$ stocks in DMSO) were added 30 min after transfection. After a further 20 h the cells were washed and fresh ligand added. After an additional 24 h, cell extracts were prepared by three cycles of freeze-thawing and assayed for β -galactosidase activity which was used to normalize the CAT assays⁵².

an ED₅₀ of about 50 μ M (Fig. 6b). Remarkably, the weak peroxisome proliferator trichloroacetic acid³² (TCA) stimulated CAT activity to a level representing 50% of that seen with 100 μ M nafenopin and with an estimated ED₅₀ of 8 mM (Fig. 6b). None of these compounds stimulates CAT activity in the absence of ER-PPAR, indicating that PPAR mediates the effects of these compounds (Fig. 5b; data not shown). In addition, some other compounds, including oestradiol (Fig. 5b, lane 8), dexamethasone (Fig. 5b, lane 16), tri-iodo-thyronine (10^{-5} M), thyroxine (10^{-5} M), retinoic acid (10^{-6} M), 25-OH-cholesterol (10^{-5} M), vitamins D (10^{-5} M) and E (10^{-6} M), dehydroepiandrosterone (DHEA, 10^{-5} M), pregnenolone 16 α carbonitrile (10^{-5} M) and phenobarbital (10^{-3} M) do not stimulate CAT activity (data not shown). Activation of ER-PPAR therefore appears to be specific for peroxisome proliferators.

Discussion

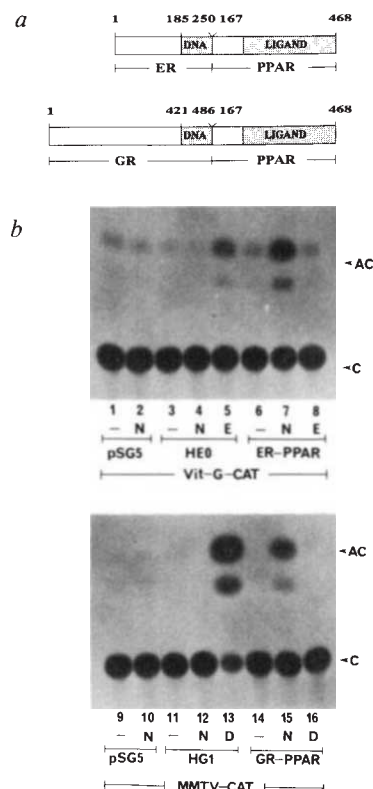
The mechanism by which peroxisome proliferators regulate gene expression and induce liver tumours in rodents is unknown. The identification of PPAR as a putative transcription factor specifically activated by peroxisome proliferators suggests that this receptor directly mediates the effects of this class of chemical. This conclusion is supported by the correlation between the rank order of peroxisome proliferators for ER-PPAR activation (Fig. 6b) and both their carcinogenicity in rodents¹ and induction of endogenous peroxisome proliferator-responsive genes³². In addition, of the mouse tissues examined, expression of PPAR (Fig. 4) is restricted to those that are responsive to peroxisome proliferators^{29,30}.

The ER-PPAR chimaera responds to many peroxisome proliferators of diverse structure (Fig. 6a). Whether any or all of these compounds can bind to PPAR is unknown, but we have been unable to demonstrate specific binding of ³H-nafenopin to ER-PPAR. This may be due in part to the apparent low

affinity of nafenopin for PPAR as judged from the dose-response curves (Fig. 6b) but also because of the large levels of endogenous binding activity observed in the cell lines tested (data not shown). Interestingly, Lalwani *et al.*^{23,24} have described a nafenopin-binding protein (PPbP) in rat liver of relative molecular mass 70,000 that is expressed at a similar level, and with a similar affinity for nafenopin, as that which we observe in nontransfected COS cells. Any relationship between PPAR and PPbP is at present unclear. However, Wy 14,643 does not bind to PPbP (refs 23, 24) whereas it strongly activates ER-PPAR (Fig. 6b), suggesting that PPAR is distinct from PPbP. Given the diversity of compounds capable of activating ER-PPAR (Fig. 6a), it remains a possibility that peroxisome proliferators act indirectly. For example, peroxisome proliferators could induce the putative endogenous ligand of PPAR or may indirectly activate the receptor for example by phosphorylation.

The physiological role and putative natural ligand of PPAR are unknown. The induction of the peroxisomal fatty-acid β -oxidation system by peroxisome proliferators may indicate a role for the receptor in providing a source of energy via fatty acids. Additional possibilities include the regulation of peroxisomal function or cholesterol metabolism, for example by increasing the conversion of cholesterol into bile acids; a step of which is performed mainly, if not exclusively, in the peroxisomes³³.

As a first step to understanding the mechanism of action of peroxisome proliferators it is important to identify peroxisome proliferator response elements (PPRE) within the regulatory regions of responsive genes. The putative DNA binding domain of PPAR contains the sequence Cys(119)-Glu-Gly-Cys-Lys-Gly (Fig. 2) important for sequence-specific DNA binding^{34–36} in receptors recognizing the DNA motif TGACC. The PPRE may therefore contain the sequence TGACC. An unusual feature of the PPAR putative DNA-binding domain is the presence of only



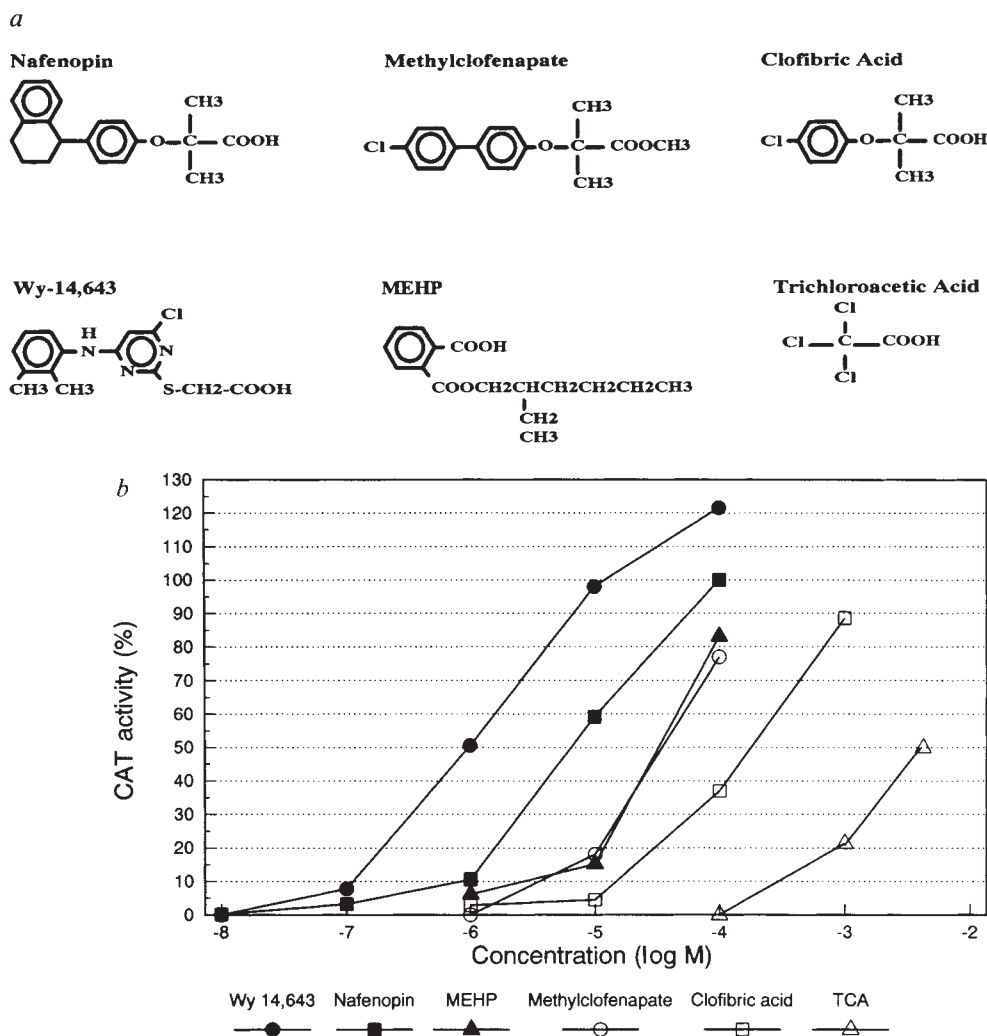


FIG. 6 Dose-response of CAT stimulation using peroxisome proliferators. **a**, Structure of the peroxisome proliferators tested. **b**, Stimulation of Vit-G-CAT by ER-PPAR in presence of peroxisome proliferators. COS1 cell transfections and CAT assays were performed as described in Fig. 5. Spots of acetylated ^{14}C -chloramphenicol were cut out from the TLC plate and quantified by liquid scintillation counting. Each point was performed in triplicate and the values vary by less than $\pm 10\%$. The results are normalized by taking the induction observed with $100\ \mu\text{M}$ nafenopin (typically 10–20-fold) as 100%.

three amino acids between Cys 139 and Cys 143 compared to the five found in all other members of this family. This region of TR β may be involved in dimerization³⁵. Possibly, PPAR binds to DNA as a monomer or, alternatively, recognizes a unique spacing of the TGACC motif.

Finally, other non-genotoxic carcinogens that seem to be liver-tumour promoters may act through additional members of the steroid hormone receptor superfamily. Indeed, such a model has been proposed for the elusive 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) receptor (see Ref. 37).

In conclusion, the identification of a peroxisome proliferator-activated receptor suggests the existence of a novel hormone-regulated system important for triglyceride and cholesterol homeostasis that may have a role in the development of liver tumours in rodents. Understanding the function of this receptor could help determine the basis for the variable response to peroxisome proliferators observed amongst species^{7,38,39} as well as aid our understanding of how these chemicals act as hypolipidaemic agents and rodent carcinogens.

Note added in proof: The nucleotide sequence of mouse PPAR has been submitted to the EMBL Data Library. □

Received 10 July; accepted 7 August 1990.

- Reddy, J. K. & Lalwani, N. D. *CRC Crit. Rev. Toxic.* **12**, 1–58 (1983).
- Reddy, J. K. & Rao, M. S. *Trends pharmac. Sci.* **7**, 438–443 (1986).
- Nemali, M. R. *et al. Toxic. appl. Pharmac.* **97**, 72–87 (1989).
- Lock, E. A., Mitchell, A. M. & Elcombe, C. R. *Rev. pharmac. Toxicol.* **29**, 145–163 (1989).
- Hess, R., Stäubli, W. & Riess, W. *Nature* **208**, 856–858 (1965).
- Reddy, J. K., Rao, M. S., Azarnoff, D. L. & Sell, S. *Cancer Res.* **39**, 152–161 (1979).
- Styles, J. A., Kelly, M., Pritchard, N. R. & Elcombe, C. R. *Carcinogenesis* **9**, 1647–1655 (1988).
- Reddy, J. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 1747–1751 (1986).
- Hijikata, M., Ishii, N., Kagamiyama, H., Osumi, T. & Hashimoto, T. *J. biol. Chem.* **262**, 8151–8158 (1987).
- Hardwick, J. P., Song, B.-J., Huberman, E. & Gonzalez, F. J. *J. biol. Chem.* **262**, 801–810 (1987).
- Kimura, S., Hardwick, J. P., Kozak, C. A. & Gonzalez, F. J. *DNA* **8**, 517–525 (1989).
- Sharma, R. K. *et al. Eur. J. Biochem.* **184**, 69–78 (1989).
- Thorp, J. M. & Waring, W. S. *Nature* **194**, 948–949 (1962).
- Havel, R. J. & Kane, J. P. *Rev. Pharmac.* **13**, 287–308 (1973).
- Sirtori, C. R., Catapano, A. & Paoletti, R. *Atherosclerosis Rev.* **2**, 113–153 (1977).
- Reddy, J. K. & Qureshi, S. A. *Br. J. Cancer* **40**, 476–482 (1979).
- Reddy, J. K., Azarnoff, D. L. & Hignite, C. E. *Nature* **283**, 397–398 (1980).
- Warren, J. R., Simmon, V. F. & Reddy, J. K. *Cancer Res.* **40**, 36–41 (1980).
- Kasai, H., Okadi, Y., Nishimura, S., Rao, M. S. & Reddy, J. K. *Cancer Res.* **49**, 2603–2605 (1989).
- Conway, J. G. *et al. Carcinogenesis* **10**, 513–519 (1989).
- Marsman, D. S., Cattley, R. C., Conway, J. G. & Popp, J. A. *Cancer Res.* **48**, 6739–6744 (1988).
- Cattley, R. C. & Popp, J. A. *Cancer Res.* **49**, 3246–3251 (1989).
- Lalwani, N. D., Fahl, W. E. & Reddy, J. K. *Biochem. biophys. Res. Commun.* **116**, 388–393 (1983).
- Lalwani, N. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 5242–5246 (1987).
- Green, S. & Chambon, P. *Trends Genet.* **4**, 309–314 (1988).
- Evans, R. M. *Science* **240**, 889–895 (1988).
- Beato, M. *Cell* **56**, 335–344 (1989).
- Kozak, M. *Nucleic Acids Res.* **15**, 8125–8148 (1984).
- Nemali, M. R. *et al. Cancer Res.* **48**, 5316–5324 (1988).
- Giacobino, J.-P. *et al. Molec. cell. Endocrin.* **61**, 217–225 (1989).
- Mitchell, A. M., Lhuguenot, J.-C., Bridges, J. W. & Elcombe, C. R. *Toxic. appl. Pharmac.* **80**, 23–32 (1985).
- Mitchell, A. M., Bridges, J. W. & Elcombe, C. R. *Archs. Toxic.* **55**, 239–246 (1984).
- Kase, B. F., Prydz, K., Björkhem, I. & Pedersen, J. I. *Biochem. biophys. Acta* **677**, 37–42 (1986).
- Mader, S., Kumar, V., de Verneuil, H. & Chambon, P. *Nature* **338**, 271–274 (1989).
- Umesono, K. & Evans, R. M. *Cell* **57**, 1139–1146 (1989).
- Danielsen, M., Hinck, L. & Ringold, G. M. *Cell* **57**, 1131–1138 (1989).
- Haggood, J., Cuthill, S., Denis, M., Poellinger, L. & Gustafsson, J. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 60–64 (1989).
- Elcombe, C. R. & Mitchell, A. M. *Environ. Hlth. Perspec.* **70**, 211–219 (1986).
- Lake, B. G., Evans, J. G., Gray, T. J. B., Körösi, S. A. & North, C. J. *Toxic. appl. Pharmac.* **99**, 148–160 (1989).
- Hollenberg, S. M. *et al. Nature* **318**, 635–641 (1985).
- Green, S. *et al. Nature* **320**, 134–139 (1986).
- Baker, A. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 3294–3298 (1988).
- Weinberger, C. *et al. Nature* **324**, 641–646 (1986).
- Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. *Nature* **330**, 444–450 (1987).

45. Giguere, V., Ong, E. S., Segui, P. & Evans, R. M. *Nature* **330**, 624–629 (1987).
 46. Lathe, R. J. *molec. Biol.* **183**, 1–12 (1985).
 47. Miyajima, N. *et al. Cell* **57**, 31–39 (1989).
 48. Feigl, G., Gram, M. & Pongs, O. *Nucleic Acids Res.* **17**, 7167–7178 (1989).
 49. Seagraves, W. A. & Hogness, D. S. *Genes Devel.* **4**, 204–219 (1990).
 50. Hamada, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 8289–8293 (1989).
 51. *Current Protocols in Molecular Biology* (Wiley, New York, 1990).
 52. Webster, N. J. G., Green, S., Jin, J. R. & Chambon, P. *Cell* **54**, 199–207 (1988).
 53. Green, S., Isseemann, I. & Scheer, E. *Nucleic Acids Res.* **16**, 369 (1988).

54. Kumar, V. *et al. Cell* **51**, 941–951 (1987).
 55. Green, S. & Chambon, P. *Nature* **325**, 75–78 (1987).
 56. Saiki, R. K. *et al. Science* **239**, 487–491 (1988).

ACKNOWLEDGEMENTS. We thank P. Chambon for plasmids, P. Bentley (Ciba Geigy) for nafenopin, D. Holland for synthesizing oligonucleotides and T. Aldridge for one of the mouse liver cDNA libraries. We also thank K. Atherton, J. Ashby, C. Elcombe, J. Tugwood and our colleagues at ICI for their support and advice.

LETTERS TO NATURE

An eclipsing millisecond pulsar in the globular cluster Terzan 5

A. G. Lyne*, R. N. Manchester†, N. D'Amico‡, L. Staveley-Smith†, S. Johnston*, J. Lim§, A. S. Fruchter||, W. M. Goss¶ & D. Frail¶

* Nuffield Radio Astronomy Laboratories, University of Manchester, Jodrell Bank, Macclesfield, Cheshire, SK11 9DL, UK

† Australia Telescope National Facility, CSIRO, Epping, NSW 2121, Australia

‡ Istituto di Fisica dell'Università di Palermo and Istituto de Radioastronomie del CNR, Bologna, Italy

§ School of Mathematics, Physics, Computing and Electronics, Macquarie University, North Ryde, NSW, Australia

|| Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA

¶ National Radio Astronomy Observatory, Socorro, New Mexico 87801, USA

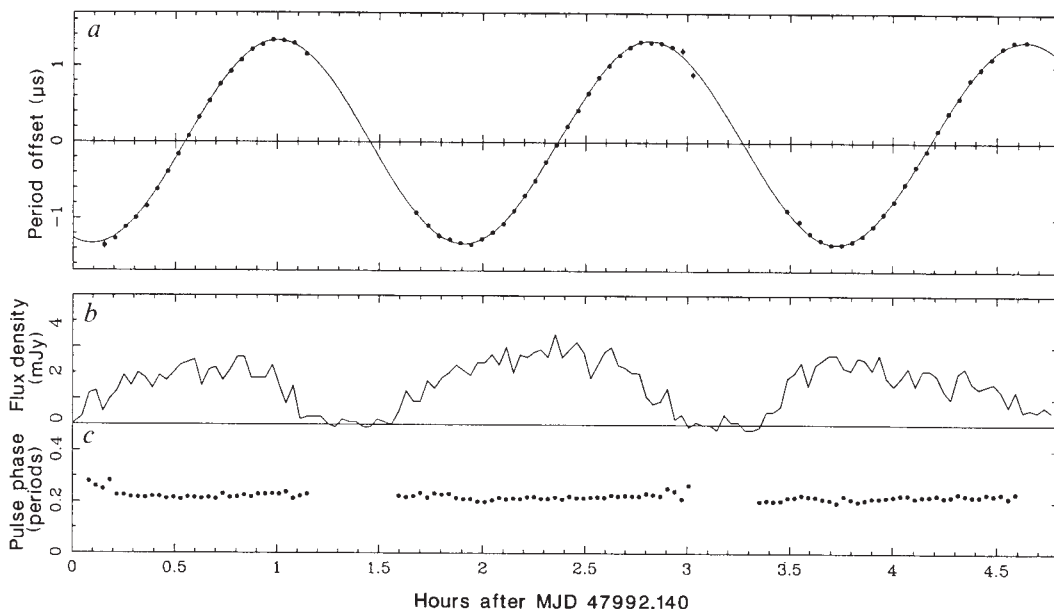
WE HAVE discovered an eclipsing binary millisecond pulsar in the globular cluster Terzan 5. This, the second known eclipsing binary pulsar after PSR1957+20, has a pulse period of 11.56 ms and a very short orbital period of 1.8 hours. In contrast to PSR1957+20, where the eclipses occupy about 10 per cent of the orbital period¹, the eclipse duration in this pulsar is very variable and never less than one-third of the orbital period. The pulsar is in a circular orbit of radius 0.11 light seconds, which implies a minimum companion mass of 0.089 solar masses, about four times the companion mass of PSR1957+20. Timing observations suggest an identification of the pulsar with a variable continuum source locate about 30 arcsec west of the cluster centre. These observations and the variable eclipse duration show that the eclipse is due to absorption or scattering in a tenuous wind which flows from the companion star. We have also detected a second pulsar in the direction of Terzan 5. This pulsar, which has a period of 442 ms,

may also be a cluster member, but is more likely to be a foreground object.

Terzan 5 is a massive globular cluster located near the Galactic Centre². It is a high-priority candidate for our pulsar search because it has a dense core³ and contains both steep-spectrum radio sources⁴ and a variable X-ray source⁵. We first detected⁶ the millisecond pulsar in observations made at 20-cm wavelength using the 64-m Parkes radio telescope in November 1989 and January 1990. The receiver used cooled field-effect-transistor preamplifiers receiving orthogonal linear polarizations and had an equivalent system flux density on cold sky of ~70 Jy. Receiver outputs were fed to two sets of filters having 64×5 MHz and 80×1 MHz channels per polarization. After summing the polarizations, the filter outputs were low-pass filtered, one-bit sampled at 0.3-ms intervals and saved on magnetic tape. At 20-cm wavelength, the telescope has a half-power beam width of 14 arcmin. Subsequent observations were made at Parkes in March and May 1990 and at Jodrell Bank using the 76-m Lovell telescope in the period March–May 1990, with receivers in the 50-cm, 30-cm, 21-cm and 18-cm bands having system equivalent flux densities of 90 Jy, 100 Jy, 45 Jy and 45 Jy, respectively. Similar filter systems having 32×250 kHz and 32×1 MHz channels per polarization were used in the Jodrell Bank observations. Outputs from these filters were sampled at intervals equal to the dispersion delay across one channel, summed with a delay of one sample interval per channel and synchronously folded at the nominal pulsar period.

Parkes data were analysed using the Convex C220 computer at the Australia Telescope National Facility. Data were initially dedispersed and then transformed using a fast-Fourier-transform routine⁷ to search for periodicities. A total of 92 candidate dispersions from 0 to $3,000 \text{ cm}^{-3} \text{ pc}$ were routinely searched for in the 20-cm observations. For Terzan 5, Fourier analysis of a 2,500-s observation revealed a signal with an amplitude about twice the noise threshold at a dispersion measure of $240 \text{ cm}^{-3} \text{ pc}$. Synchronous folding of the data at the nominal period showed

FIG. 1 Variations in *a*, the Solar System barycentre period and *b*, the pulsed flux density for the binary pulsar in Terzan 5 from data recorded at Jodrell Bank at a wavelength of 18 cm on 11 April 1990. Orbital phase is with respect to the ascending node (so that the pulsar is behind the companion at phase 0.25). The central period for *a* is that given in Table 1. *c*, The observed pulse phase for the same observation after compensation for both the binary motion of the pulsar and the motion of the observatory around the Sun.



that the apparent pulsar period was varying rapidly during the observation, implying that the pulsar was a member of a short-period binary system. Furthermore, only the ascending portion of the period curve was seen, suggesting that the pulsar was in eclipse for about half the orbital period. These results were confirmed in further observations made at Parkes and at Jodrell Bank. The observed period and amplitude variations of this pulsar, which we have called PSR1744–24A, are illustrated in Fig. 1*a, b*.

Images of the 20-cm continuum radiation from the direction of Terzan 5 made at the Very Large Array (VLA) in June and September 1989 and March and April 1990, show several point sources with amplitudes ≤ 2 mJy within a few arcmin of the cluster centre. Analyses of pulse time of arrivals from both Jodrell Bank and Parkes for the interval 2 March–19 May 1990 are consistent with the identification of PSR1744–24A with one of these sources, which is located ~ 30 arcsec west of the cluster centre and highly variable in amplitude. Table 1 lists the observed parameters of the pulsar. Timing parameters are from a fit with the eccentricity and longitude of periastron set to zero; the limit on the eccentricity was determined in a separate fit. The timing parameters were referred to the Solar System barycentre using the DE200 ephemeris⁸ and observations during and adjacent to eclipses were omitted from the fit. There is no evidence of an interpulse. The estimated distance³ to the cluster is 7.1 kpc, and from the dispersion measure, the mean interstellar electron density in the path to the pulsar is 0.034 cm^{-3} . This value is typical for pulsars lying close to the galactic plane⁹ and hence quite consistent with the pulsar being a cluster member. The pulsar lies well outside the cluster core, however, which has an estimated radius³ of only about 3 arcsec.

Variations in pulse amplitude and phase on different days are illustrated in Fig. 2. In contrast to PSR1957+20, where the eclipse duration at a given frequency is relatively stable and symmetric about orbital phase 0.25 (ref. 10), the eclipses of PSR1744–24A are extremely variable. Their duration ranges from a minimum of about one-third of the orbital period to total; on roughly one out of five days the pulsar was undetected for the entire six-hour observation at Jodrell Bank. The orbital phase at eclipse emersion is much more stable than that at immersion and is generally between orbital phases of 0.4 and

TABLE 1 Observed parameters of PSR 1744–24A

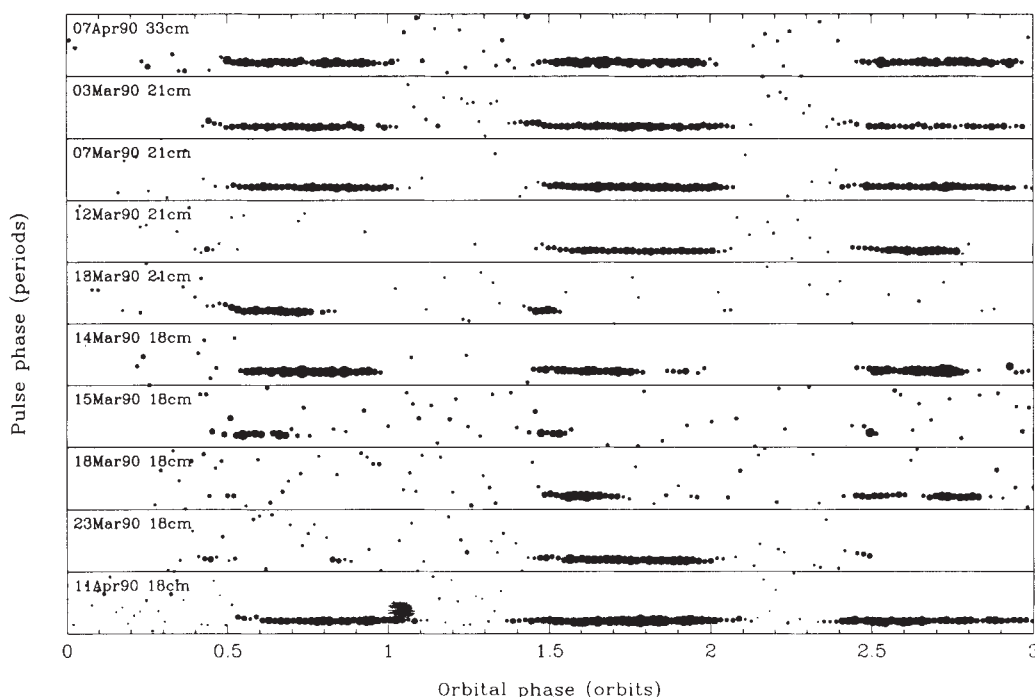
Right ascension (B1950)	17 ^h 44 ^m 57.70 ^s
Declination (B1950)	–24° 45′ 38.1″
Galactic longitude	3.84°
Galactic latitude	+1.70°
Barycentric period (ms)	11.563 148 390 6 (12)
Period derivative	$(-5 \pm 9) \times 10^{-20}$
MJD of period	47837.0000
Orbital period (d)	0.075 646 114 (4)
Projected semi-major axis (s)	0.119 67 (2)
Eccentricity	$< 10^{-3}$
MJD of ascending node	47 953.299 686 (2)
Dispersion measure ($\text{cm}^{-3} \text{ pc}$)	242.0 (2)
r.m.s. residual (μs)	83
Mass function (solar masses)	3.215×10^{-4}
Mean flux density at 50 cm (mJy)	5
Mean flux density at 20 cm (mJy)	2.5
Pulse width (ms)	0.7

Numbers in parentheses are the 2σ errors in the last significant figures. The uncertainty on the position, determined at the VLA, is 0.4 arcsec. MJD, mean Julian date. Quoted flux densities are averages for intervals when the pulsar was not in eclipse. The pulse width (full width at half maximum) has been corrected for instrumental broadening.

0.5. Although the great variations in eclipse duration at a given frequency complicate the analysis, there is evidence that eclipses are longer at lower frequencies, as is observed for PSR1957+20 (ref. 10). The pulsar was undetected in VLA observations at 327 MHz with an upper limit on the flux density of 10 mJy.

Figures 1*a* and 2 show that the pulse arrival times observed outside of the eclipses are mostly well fitted by the timing model. The decay and rise of pulse amplitude is gradual and during this interval the pulse phase is generally stable with any excess delay being less than a few hundred microseconds. But there are occasional significant increases in pulse delay near the eclipse, especially on the emersion side. These excess delays are very similar in form to, although larger than, those observed in PSR1957+20 (ref. 10). Intervals of phase-coherent signals apparently occur during some eclipses. If genuine, they indicate propagation delays of at least several milliseconds during the eclipse.

FIG. 2 Observed pulse phase, after compensation for both the binary motion of the pulsar and the motion of the observatory around the Sun, plotted as a function of orbital phase for three consecutive orbits on each of 10 days from data recorded at Jodrell Bank. For each observation, one complete period of pulse phase is shown. Signal amplitudes are represented by the size of the dot, with dot diameter being proportional to the amplitude above a cutoff. The observation date and wavelength band is indicated in each panel. Although most eclipses last for about half of the orbital period, their duration varies greatly from day to day. Successive eclipses tend to be similar, however, showing that variations in the eclipsing medium have a timescale of the order of half a day.



If we assume that the pulsar is a neutron star of mass $1.4 M_{\odot}$, the minimum companion mass is $0.089 M_{\odot}$. The presence of eclipses shows that the perpendicular to the orbit plane is not at a small angle to the line of sight, so the actual companion mass must be close to this value. There is no doubt that the eclipses are due to a wind emanating from the companion, as in PSR1957+20 (ref. 10). The observations that eclipses are often more than half the orbital period and that the pulse amplitude varies continuously (Fig. 1) show that the eclipsing medium often envelopes the pulsar.

The wind in this system may be ablated from a degenerate dwarf companion, as is believed to be the case for PSR1957+20 (refs 11,12). The energy flux from the pulsar at the companion (assuming isotropic radiation) is $<2 \times 10^{10} \text{ erg s}^{-1} \text{ cm}^{-2}$, however, at least a factor of 15 less than for PSR1957+20 (ref. 10). This, and the great variations in eclipse duration, suggest that the wind may result from material overflowing the Roche lobe of the companion. The estimated radius of a main-sequence star with a mass equal to the companion minimum mass^{13,14} is only $\sim 10\%$ smaller than the Roche-lobe radius of the companion, so flow of gas into the Roche lobe of the pulsar is quite probable. Because of the binary motion, this gas would tend to trail the pulsar as it fell toward it, possibly accounting for the observed distribution of eclipse phases.

The detection of phase-coherent pulses well into most eclipses with only small pulse delays indicates that these eclipses are most likely due to absorption or wide-angle scattering. For the deepest parts of the eclipse, propagation delays may be several milliseconds or more. If these delays are dispersive, they correspond to a change of several $\text{cm}^{-3} \text{ pc}$ in the dispersion measure, two orders of magnitude greater than those in PSR1957+20 (ref. 1). If we further assume that the effective path is of the order of the orbital radius of the companion, then electron densities of $\sim 10^8 \text{ cm}^{-3}$ are implied. The plasma frequency of such a medium is $\sim 100 \text{ MHz}$, suggesting that refractive effects are not important unless the medium is very clumpy. Free-free absorption, as discussed for PSR1957+20 by Rasio *et al.*¹⁵, is the most probable eclipse mechanism. For a uniform wind of the depth and density assumed above and an optical depth of ~ 1 , the electron temperature must be $\sim 10^5 \text{ K}$; this is high but perhaps not unreasonable for an accretion flow. If the eclipses are due to absorption or wide-angle scattering in the wind, we would expect similar eclipses in the VLA source. This supports the identification of the pulsar with the variable source discussed above.

Analysis of 50-cm data recorded at Parkes in December 1989 has also revealed a slower pulsar, PSR1744–24B, in the direction of Terzan 5. This pulsar has a heliocentric period of $442.839 \pm 0.001 \text{ ms}$ and a dispersion measure of $205 \pm 5 \text{ cm}^{-3} \text{ pc}$. It was also detected in Parkes 20-cm data with a mean flux density of $\sim 0.3 \text{ mJy}$. There are several sources in the VLA image at about this flux density and, at this stage, it is not possible to identify which of them is this pulsar. The dispersion measure is significantly different from that of PSR1744–24A. We have argued previously¹⁶ that dispersion-measure contributions of the order of $35 \text{ cm}^{-3} \text{ pc}$ are unlikely to occur in a globular cluster. Such a column density of electrons would require trapping of wind gas, which is unlikely in the globular cluster environment. It is likely that PSR1744–24B is a foreground pulsar. The density of such pulsars at this flux density and in this direction is expected to be about two per square degree¹⁷, and so the probability of finding one within the 20-cm beam is $\sim 10\%$. $\square$

5. Makishima, K. *et al. Astrophys. J.* **247**, L23–L25 (1981).
6. Lyne, A. G. *et al. IAU Circ. No.* 4974 (1990).
7. Fraser, D. *ACM Trans. Math. Software* **4**, 500–517 (1979).
8. Standish, E. M. *Astr. Astrophys.* **114**, 297–302 (1982).
9. Lyne, A. G., Manchester, R. N. & Taylor, J. H. *Mon. Not. R. astr. Soc.* **213**, 613–639 (1985).
10. Fruchter, A. S. *et al. Astrophys. J.* **351**, 642–650 (1990).
11. Phinney, E. S., Evans, C. R., Blandford, R. D. & Kulkarni, S. R. *Nature* **333**, 832–834 (1988).
12. Kluzniak, W., Ruderman, M., Shaham, J. & Tavani, M. *Nature* **334**, 225–227 (1988).
13. Popper, D. M. *A. Rev. Astr. Astrophys.* **18**, 115–164 (1980).
14. Biermann, P. *et al. Astrophys. J.* **293**, 303–320 (1985).
15. Rasio, F. A., Shapiro, S. L. & Teukolsky, S. A. *Astrophys. J.* **342**, 934–939 (1989).
16. Manchester, R. N. *et al. Nature* **345**, 598–600 (1990).
17. Clifton, T. R. & Lyne, A. G. *Nature* **320**, 43–45 (1987).

ACKNOWLEDGEMENTS. We thank M. Bailes, E. van den Heuvel and J. Liebert for discussions, F. Rasio for comments and M. Bailes for help with the Jodrell Bank observations. The Australia Telescope National Facility is operated in association with the Division of Radiophysics by CSIRO and the National Radio Astronomy Observatory is operated by Associated Universities Inc. under cooperative agreement with the NSF. N.D.A. acknowledges financial support from CNR and Ministero Pubblica Istruzione.

Raman scattering as a diagnostic tool in the massive X-ray binary 4U1700–37

L. Kaper, G. Hammerschlag–Hensberge & R. J. Takens

Anton Pannekoek Astronomical Institute, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

HD153919, an extreme Of star (spectral type O6.5Iaf⁺; ref. 1) has been identified as the optical counterpart of the 3.411-day eclipsing massive X-ray binary 4U1700–37 (ref. 2) by spectroscopic^{3–5} and photometric⁶ studies which revealed periodic optical variations in phase with the X-ray period. Here we report that the ultraviolet spectra taken with the International Ultraviolet Explorer satellite also vary in phase with the X-ray emission. In an interval of $\sim 200 \text{ \AA}$ centred on the subordinate He II line at $1,640 \text{ \AA}$, emission lines $2\text{--}3 \text{ \AA}$ wide appear and disappear, with maximum flux at binary phase 0.5 (when the neutron star is in front of the O supergiant) and minimum at phase 0.8. We suggest that these variable lines are due to Raman scattering of extreme ultraviolet photons from the X-ray source by He II ions. These Raman lines can thus be used to reconstruct a part of the unobserved extreme ultraviolet spectrum in the He II scattering zone.

The ultraviolet spectrum of HD153919 shows saturated P Cygni profiles^{7,8} of the resonance lines of N V at $1,240 \text{ \AA}$, Si IV at $1,400 \text{ \AA}$ and C IV at $1,550 \text{ \AA}$, indicative of a strong, dense stellar wind. The mass loss rate is estimated to be $6 \times 10^{-6} M_{\odot} \text{ yr}^{-1}$ and the terminal velocity of the wind is $\sim 2,200 \text{ km s}^{-1}$ (ref. 9).

The X-ray light curve of the system exhibits strong flaring activity and a very long and variable eclipse duration^{10,11}. X-ray pulsations have not been detected¹², but the similarity of the X-ray spectrum of 1700–37 to those of other X-ray pulsars¹³

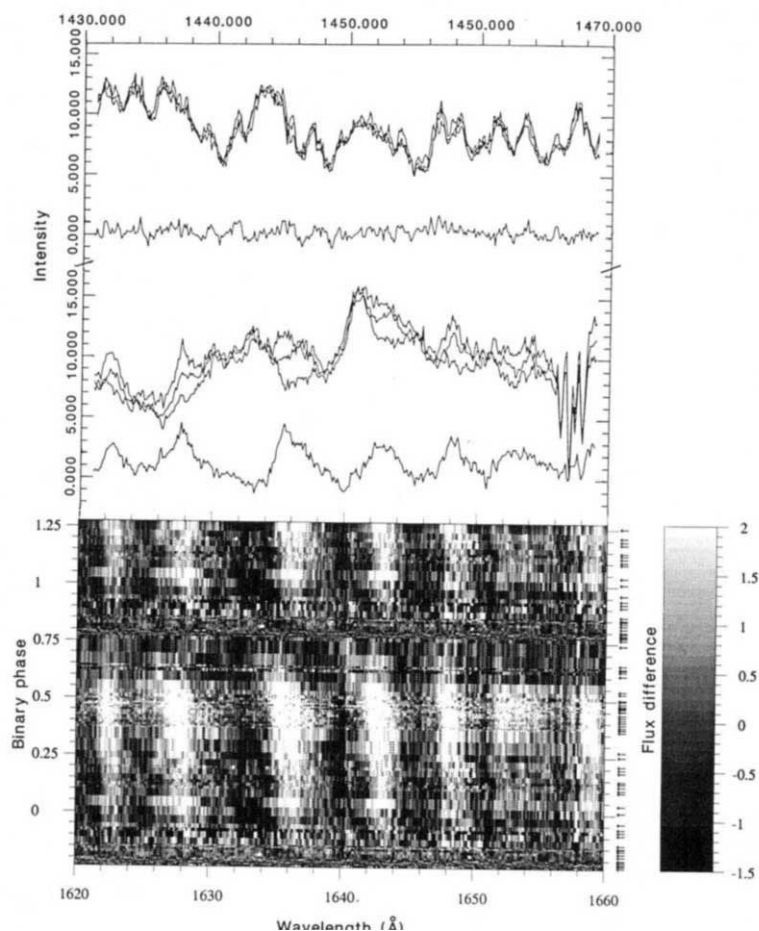
TABLE 1 Observed Raman lines and input EUV emission lines

$\lambda_{\text{Raman}} (\text{\AA})$	$\lambda_{\text{EUV}} (\text{\AA})$	$\lambda_{\text{Raman}} (\text{\AA})$	$\lambda_{\text{EUV}} (\text{\AA})$	$\lambda_{\text{Raman}} (\text{\AA})$	$\lambda_{\text{EUV}} (\text{\AA})$
1,522	253.24	1,627	256.00	1,721	258.20
1,529	253.43	1,636	256.21	1,727	258.35
1,558	254.21	1,643	256.38	1,743	258.70
1,565	254.40	1,648	256.50	1,749	258.83
1,572	254.58	1,653	256.62	1,757	259.00
1,581	254.83	1,660	256.79	1,762	259.11
1,587	254.97	1,668	256.98	1,788	259.66
1,593	255.13	1,678	257.21	1,806	260.04
1,598	255.26	1,683	257.33	1,815	260.23
1,606	255.46	1,694	257.59	1,824	260.41
1,612	255.62	1,707	257.90	1,831	260.55
1,618	255.76	1,713	258.02	1,850	260.93
1,622	255.87				

Received 8 June; accepted 14 August 1990.

1. Fruchter, A. S., Stinebring, D. R. & Taylor, J. H. *Nature* **333**, 237–239 (1988).
2. Terzan, A. *Astr. Astrophys.* **12**, 477–481 (1971).
3. Webbink, R. F. *Dynamics of Star Clusters, IAU Symp. 113* (eds Goodman, J. & Hut, P.) 541–577 (Reidel, Dordrecht, 1985).
4. Fruchter, A. S. & Goss, W. M. *Bull. Am. astr. Soc.* **21**, 1204 (1990).

FIG. 1 The upper part of the figure shows three averaged spectra (around $\phi=0.15$, $\phi=0.45$ and $\phi=0.80$) for two spectral regions centred around 1,450 Å (top) and the subordinate He II line at 1,640 Å (bottom). Under each display the difference spectrum of the $\phi=0.45$ and $\phi=0.80$ spectra is shown. The lower part of the figure shows a grey-scale image of all 60 archival IUE spectra in the same wavelength region, arranged according to orbital phase. Here, the averaged spectrum around $\phi=0.8$ was subtracted from all spectra. The grey scale represents the flux difference: white corresponds to maximum flux difference.



indicates that the X-ray source is an accreting neutron star. Also the optical mass function⁴ favours a neutron-star companion. The average (uneclipsed) X-ray luminosity of $\sim 10^{36}$ erg s⁻¹ can be explained by accretion of stellar-wind material from the supergiant onto the neutron star. The X-ray light curve often shows absorption dips around phase 0.75, accompanied by an overall increase in spectral hardness ratio after phase 0.7 (ref. 14), which indicates the presence of material trailing the neutron star in its orbit. Observations of the time-dependent behaviour of the He I 5,876 Å and the H α line^{3,5} support this interpretation.

We reduced all existing high-resolution ultraviolet spectra of HD153919 obtained by the International Ultraviolet Explorer (IUE) satellite, using the IUEDR software package on the Starlink network¹⁵, which resulted in a total of 60 SWP spectra (short wavelength: 1,150–1,950 Å) spread over 9 years (1978–86). The (old) small-aperture spectra were corrected for flux losses during observation, and wavelength calibration was performed by shifting interstellar lines towards their rest wavelength (for more details, see ref. 8).

Figure 1 shows the region centred around the subordinate He II 1,640 Å line for three averaged spectra around $\phi=0.15$, $\phi=0.45$ and $\phi=0.80$. The three spectra are quite different. A number of 2–3 Å wide (weak) emission lines are present around phase 0.45 and absent around phase 0.80. For comparison, the same spectra are shown in an interval around 1,450 Å: here no differences are found. A grey-scale representation of all 60 spectra around 1,640 Å, sorted by binary phase, demonstrates that the variations are a function of orbital phase. To represent more clearly the differences between the individual spectra, we subtracted the averaged $\phi=0.8$ spectrum from all spectra. The emission lines are variable in flux, with a maximum around $\phi=0.5$ and a minimum at $\phi=0.8$. There is an abrupt decrease

in flux after the maximum is reached. Figure 2a shows the flux averaged over all observable lines as a function of orbital phase. This picture remarkably resembles the variation with orbital phase of the X-ray spectral hardness¹⁴, which suggests that the obscuration of the emission lines is connected with the rise in N_H column density during late phases¹¹.

To obtain information on the region of formation of these broad emission lines, we measured their central wavelengths as a function of binary phase. In Fig. 2b the velocity shift of the emission lines is plotted as a function of orbital phase. The data were fit (least squares) with a sinusoidal function of the form $\Delta v = \alpha + \beta \sin(2\pi\phi - \gamma)$, where Δv is the mean velocity difference. The best fit (with 2 σ errors) obtained was

$$\Delta v = (26 \pm 8) \sin(2\pi[\phi - (0.87 \pm 0.04)]) \text{ km s}^{-1} \quad (1)$$

The parameter α is 'uninteresting' in the sense of Avni (ref. 16, section II). The binary nature of the system is indicated by the Doppler shift of the emission lines. The velocity amplitude differs per line; the value of β is an average over all lines. The velocity variation is not in phase with either the O star or the neutron star. To decide precisely where the emission lines are formed, a detailed model of the system is needed to derive conclusions from the phase difference and velocity information.

Analysis leads to the following results: (1) the emission lines are broad: 2–3 Å wide; (2) they are concentrated in a ~ 200 Å region centred around He II 1,640 Å. Other parts of the 1,150–1,950 Å region covered by the SWP camera do not show the above-mentioned variations (Fig. 3); (3) the lines vary in flux; they reach a maximum at phase 0.5 (where the neutron star is in front of the O star), and decrease abruptly to a minimum near phase 0.8; (4) the central wavelength of the emission lines is Doppler shifted according to binary phase.

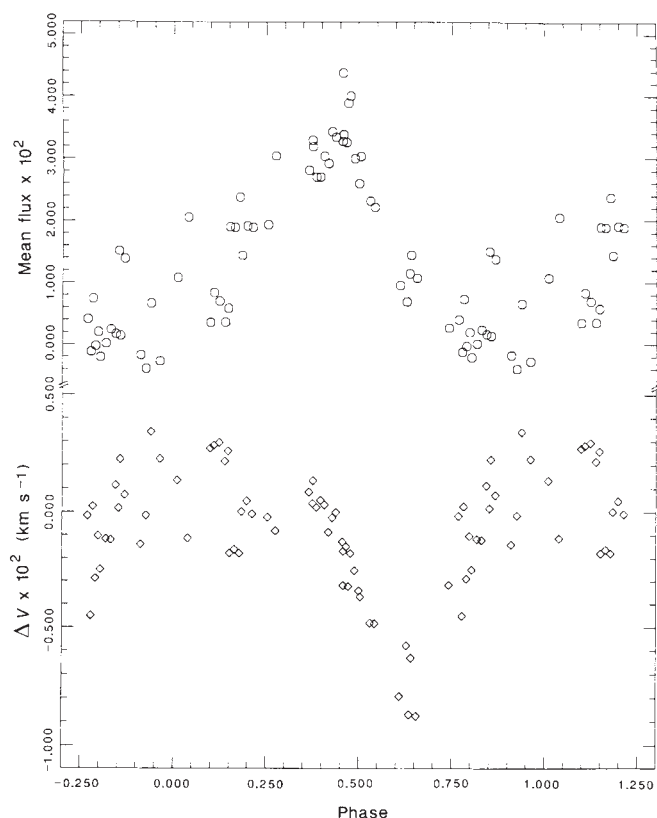


FIG. 2 *a*, Flux, averaged over all observable emission lines, shown as a function of orbital phase. There is an abrupt decrease in flux towards late phases. *b*, Velocity shift of the emission lines with respect to their mean central wavelength plotted against orbital phase. There is a small, but significant, phase shift with respect to the velocity curve of the O star.

We suggest Raman scattering (see, for example, ref. 17) as a mechanism for the formation of the observed variable emission lines. This process is like fluorescent scattering, with the difference that the intermediate state (excited atom) in Raman scattering, does not correspond to an eigenstate of the atom. It can be described as lying in the far wing of a nearby energy level with suitable quantum numbers, so the Lorentz profile of this energy level appears in the cross-section; this makes the

cross-section for Raman scattering quite small. To compensate for this, the input must be a strong emission line (assumed frequency ν_i). This line should be rather close to a strong line (frequency ν_0) of a very abundant ion. In practice, the latter line should arise from the ground state of this ion. After absorption of a photon at frequency ν_0 , the ion must have an alternative way back, starting with emission of a photon of frequency ν_1 ; when $\nu_1 = \nu_0$ the process is called resonant scattering (in the far wing). A Raman-scattered photon ν_f has frequency

$$\nu_f = \nu_i + \nu_1 - \nu_0 \quad (2)$$

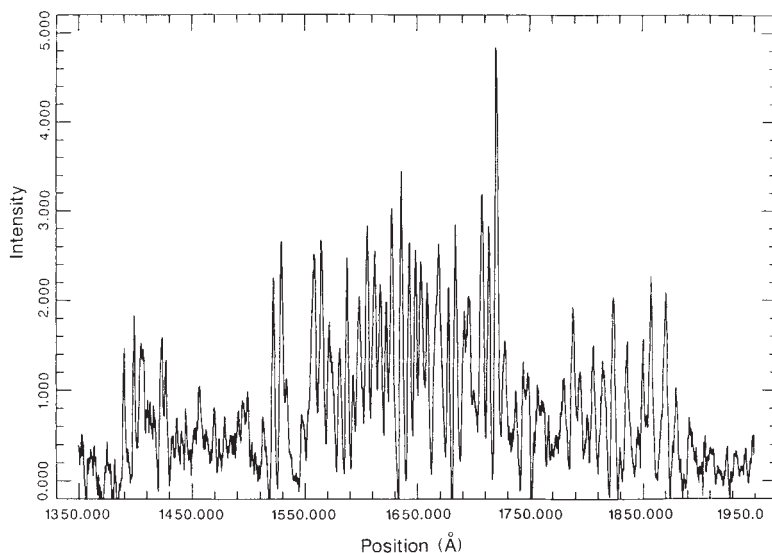
Because the wing of the upper level of ν_0 is the intermediate state, the cross-section for Raman scattering varies as $(\nu_i - \nu_0)^{-2}$, and only lines with frequency ν_i close to ν_0 are good candidates. Then, according to equation (2), ν_f is close to ν_1 . A resonance line, like Lyman α , cannot provide the frequency ν_0 , because there is no $\nu_1 \neq \nu_0$. Emission lines close to Lyman β , however, can be Raman scattered to the spectral region around Balmer α . The large linewidth is characteristic for this scattering: according to equation (2) we have $\Delta\nu_f = \Delta\nu_i$; expressed in the usual way this gives

$$\frac{\Delta\nu_f}{\nu_f} = \frac{\nu_i}{\nu_f} \frac{\Delta\nu_i}{\nu_i} \quad (3)$$

This width is enhanced by a factor ν_i/ν_f (in our case a factor of six, see below) compared to the width of the line of frequency ν_i . Both the large width, and the lack of any identification, point to Raman scattering as underlying mechanism for the emission lines. The abundant ion involved (corresponding to ν_0 and ν_1) is easily identified as singly ionized helium, which is very abundant in the stellar wind of an O6.5Iaf⁺ star. In the SWP high-resolution spectra the emission lines are concentrated in an interval between 1,525 and 1,750 Å. This is symmetric around 1,640 Å, the 'Balmer α ' line of He II. So the EUV 'input' lines should be found close to 256 Å, the He II 'Lyman β ' line.

The wavelengths of the Raman-scattered emission lines can be used to calculate the wavelengths of the corresponding EUV lines (see Table 1). It is, however, not trivial to identify the emission lines in the EUV region. Except for the Sun, a stellar EUV spectrum below 300 Å is never observed in high resolution. In solar-flare spectra^{18,19}, many emission lines from highly ionized species, like Si X, Fe XXII, S X, are found in this region. But these spectra are not of high enough resolution to resolve all lines (also, the physical conditions may be quite different from those around a neutron star). The most recently calculated (but incomplete) X-ray spectra in this wavelength region²⁰ of optically thin plasmas contain much less than the ~35 EUV

FIG. 3 A smoothed difference spectrum (of the averaged $\phi = 0.5$ and $\phi = 0.8$ spectra) from 1,400 to 1,900 Å showing the region of emission lines centred around He II 1,640 Å. The Lorentz shape of the upper energy level of He II 256 Å can be deduced from this figure.



lines listed in Table 1. Neither theoretical^{21,22} nor laboratory²³ EUV spectra contain enough information to make unique identifications.

The flux variation of these Raman lines is probably connected with the obscuration of the neutron star by dense filaments in the disrupted stellar wind²⁴. The diagnostic properties of Raman scattering may profit from high-resolution ultraviolet observations with Hubble Space Telescope in the near future. We expect that observations with the Hubble telescope of other massive X-ray binaries with O-star companions, such as Cen X-3, Cyg X-1 and LMC X-4 (which are too faint for IUE), can help to solve the identification problem. In particular, the lower metallicity in the Large Magellanic Cloud may result in emission-line fluxes of different strength. Further, we expect progress in theoretical spectra in the next few years (J. C. Raymond, personal communication) which may make Raman scattering a powerful diagnostic tool to study parts of the unobserved EUV spectrum. □

Received 4 June; accepted 24 August 1990.

- Walborn, N. R. *Astr. J.* **78**, 1067–1073 (1973).
- Jones, C. *et al.* *Astrophys. J.* **181**, L43–L48 (1973).
- Conti, P. S. & Cowley, A. P. *Astrophys. J.* **200**, 133–144 (1975).
- Hammerschlag-Hensberge, G. *Astr. Astrophys.* **64**, 399–405 (1978).
- Fahlman, G. G. & Walker, G. A. H. *Astrophys. J.* **240**, 169–179 (1980).
- Van Paradijs, J., Hammerschlag-Hensberge, G. & Zuiderwijk, E. J. *Astr. Astrophys. Suppl. Ser.* **31**, 189–198 (1978).
- Dupree, A. K. *et al.* *Nature* **275**, 400–403 (1978).
- Hammerschlag-Hensberge, G., Howarth, I. D. & Kallman, T. R. *Astrophys. J.* **352**, 698–708 (1990).
- Howarth, I. D. & Prinja, R. K. *Astrophys. J. Suppl. Ser.* **69**, 527–592 (1989).
- Branduardi, G., Mason, K. O. & Sanford, P. W. *Mon. Not. R. astr. Soc.* **185**, 137–142 (1978).
- Haberl, F., White, N. E. & Kallman, T. R. *Astrophys. J.* **343**, 409–425 (1989).
- Gottwald, M., White, N. E. & Stella, L. *Mon. Not. R. astr. Soc.* **222**, 21P–25P (1986).
- White, N. E., Kallman, T. R. & Swank, J. H. *Astrophys. J.* **269**, 264–272 (1983).
- Mason, K. O., Branduardi, G. & Sanford, P. W. *Astrophys. J.* **203**, L29–L33 (1976).
- Giddings, J. R. *ESA IUE Newsletter* **17**, 53–59 (1983).
- Avni, Y. *Astrophys. J.* **210**, 642–646 (1976).
- Nussbaumer, H., Schmid, H. M. & Vogel, M. *Astr. Astrophys.* **211**, L27–L30 (1989).
- Behring, W. E., Cohen, L., Feldman, U. & Doschek, G. A. *Astrophys. J.* **203**, 521–527 (1976).
- Dere, K. P. *Astrophys. J.* **221**, 1062–1067 (1978).
- MacFarlane, J. J. & Cassinelli, J. P. *Astrophys. J.* **347**, 1090–1099 (1989).
- Mewe, R., Gronenschild, E. H. B. M. & Van den Oord, G. H. J. *Astr. Astrophys. Suppl. Ser.* **62**, 197–254 (1985).
- Raymond, J. C. & Smith, B. W. *Astrophys. J. Suppl. Ser.* **35**, 419–439 (1977).
- Kelly, R. L. *J. Phys. Chem. Ref. Data* **16**, Suppl. 1 (1987).
- Blondin, J. M., Kallman, T. R., Fryxell, B. A. & Taam, R. E. *Astrophys. J.* (in the press).

ACKNOWLEDGEMENTS. We thank H. F. Henrichs for providing the software for the lower part of Fig. 1 and J. J. MacFarlane, J. C. Raymond and M. Vogel for discussions. We also thank J. van Paradijs and W. Penninx for comments.

Isotope mass fractionation during evaporation of Mg_2SiO_4

Andrew M. Davis*, Akihiko Hashimoto‡, Robert N. Clayton*†‡ & Toshiko K. Mayeda*

* Enrico Fermi Institute, † Department of Chemistry, and ‡ Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois 60637, USA
§ Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA

CALCIUM- and aluminium-rich refractory inclusions (CAIs) in carbonaceous chondrites can contain Mg, Si and Cr with isotopic compositions that are significantly mass-fractionated, up to a few per cent per a.m.u., relative to normal (terrestrial) values^{1,2}. The oxygen isotopic composition of so-called FUN (fractionated and unknown nuclear effect) CAIs is also interpreted to be highly mass-fractionated from a postulated ¹⁶O-rich reservoir³. Isotope fractionation between vapour and solids or liquids is likely to be responsible for these effects. Because of the large mass fractionations observed, it has been proposed that they are kinetic isotope effects produced by non-equilibrium evaporation and/or condensation processes that occurred in the primordial solar nebula³. To

test this idea, we have partially evaporated synthetic forsterite (Mg_2SiO_4) in vacuum for various durations and at different temperatures. The residual charges obtained when molten Mg_2SiO_4 is evaporated to 12% of its initial mass are enriched in heavy isotopes by ~20, ~30 and ~15‰ a.m.u.⁻¹ for O, Mg and Si, respectively, whereas solid forsterite evaporated to a similar residual mass fraction shows negligible fractionations (<1‰ a.m.u.⁻¹). These results imply that CAIs must have been at least partially molten if the observed large mass fractionation effects were caused by evaporation.

Isotope mass fractionation during evaporation of rocks and minerals has been investigated previously. Molini-Velsko *et al.*⁴ reported Si isotope analyses of evaporation residues of a basalt and two carbonaceous chondrites produced by heating in a solar furnace. The residues were found to be enriched in the heavy isotopes of Si, but only by 1.1–1.8‰ a.m.u.⁻¹ after 80% of the sample had been evaporated. Esat *et al.*⁵ evaporated natural diopsidic pyroxene in a carbon boat in vacuum using an induction heater. They found that both the residue and the material that condensed in the vacuum apparatus had isotopically mass fractionated Mg, with the residue enriched in heavy isotopes by up to 27‰ a.m.u.⁻¹ and the condensate depleted by as much as 14‰ a.m.u.⁻¹. (They also reported that the isotopically heavy residues contained up to 1% deficit in ²⁶Mg and the condensates contained up to a 2% excess in ²⁶Mg relative to their calculated mass fractionated compositions.) Those evaporation experiments did not produce true kinetic isotope fractionation, because they did not fulfill the condition of free evaporation⁶. Each study was done on a single element, so the relationship between

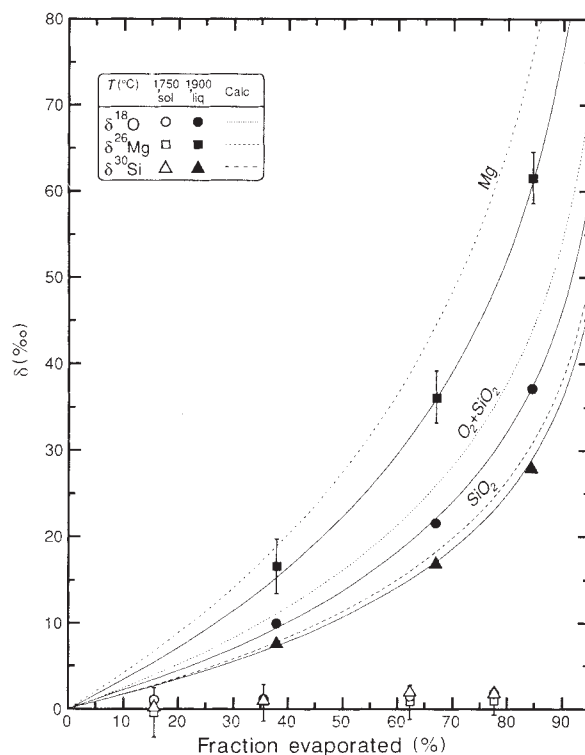


FIG. 1 Isotopic compositions of Mg_2SiO_4 evaporation residues, expressed as $\delta^{18}\text{O}$, $\delta^{26}\text{Mg}$ and $\delta^{30}\text{Si}$ relative to starting material. The solid curves are the best-fit Rayleigh fractionation curves for each element for experiments run at 1,900 °C. Experiments run at 1,940 and 2,050 °C, which lie slightly below the 1,900 °C Rayleigh fractionation curves, are not shown for clarity (see Table 1). The dashed curves represent theoretical Rayleigh fractionation for each element, obtained with the fractionation factors calculated from the inverse square roots of the masses of the evaporating species. The species are indicated above the curves. The measured isotopic compositions are close to, but always below, the theoretical curves.

TABLE 1 Isotopic compositions of Mg_2SiO_4 residue samples

T ($^{\circ}\text{C}$)	% Evaporated	Time (min)	$\delta^{17}\text{O}$	$\delta^{18}\text{O}$	$\delta^{25}\text{Mg}$	$\delta^{26}\text{Mg}$	$\delta^{29}\text{Si}$	$\delta^{30}\text{Si}$
1,750	15.6	60	0.54	1.04		-0.41		0.24
1,750	35.6	180	0.59	1.12		0.72		0.98
1,750	62.4	300	1.08	1.48		0.78		1.91
1,750	77.7	370	0.94	1.80		0.89		1.90
1,900	38.0	20	4.99	9.89	7.52	16.61	3.78	7.54
1,900	67.0	38	10.94	21.50	16.95	36.02	8.52	16.82
1,900	84.4	64.5	18.86	37.06	30.31	61.39	14.02	27.82
1,940	54.2	11	5.88	11.81	9.89	21.87	4.65	9.40
2,050	42.3	4.0	4.82	9.62	6.09	13.20	4.20	7.63
2,050	67.4	9.2	11.30	22.37	16.60	34.14	7.05	13.87
2,050	87.8	15.2	21.41	42.55	31.77	63.74	15.19	29.51
$\pm 2\sigma$			0.05	0.05	1.79	3.08	0.05	0.05

$\delta^{17}\text{O}(\text{‰}) = \{[(^{17}\text{O}/^{16}\text{O})_{\text{residue}} / (^{17}\text{O}/^{16}\text{O})_{\text{start}}] - 1\} \times 1,000$, and other isotopic compositions defined similarly.

isotope mass fractionations in different elements could not be investigated. We used olivine as a starting material because it is the most abundant mineral in stony meteorites and involatile planetary matter, and it contains all of the major elements in question. Another advantage is that Mg_2SiO_4 evaporates in stoichiometric elemental proportions, either from the solid or liquid state, and its evaporation mechanism is understood⁷.

The experimental apparatus and technique are described in detail elsewhere⁷. Here Mg_2SiO_4 was partially evaporated from the solid state at 1,750 $^{\circ}\text{C}$ for 1–6 h, and from the molten state at 1,900, 1,940 and 2,940 $^{\circ}\text{C}$ for 4–64 min. Experimental run products lost 15.6–87.8 wt% by evaporation. For each evaporation residue, O and Si isotopes were measured in bulk by gas-source mass spectrometry and Mg isotopes were measured on a polished section by ion-microprobe mass spectrometry.

Table 1 shows that remarkably little isotope fractionation of O, Mg and Si was produced in the solid forsterite residues at 1,750 $^{\circ}\text{C}$, despite evaporation of 77.7% of the starting amount in the most extreme case. In contrast, evaporation of Mg_2SiO_4 above the melting point of 1,890 $^{\circ}\text{C}$ produced a residue strongly enriched in the heavy isotopes of all elements. The degree of enrichment in heavy isotopes increases with increasing degree of evaporation. Our experiments provide independent fractionation factors for two isotope pairs for each element (such as $^{17}\text{O}/^{16}\text{O}$ and $^{18}\text{O}/^{16}\text{O}$), and thus provide tests for the fractionation laws relating the magnitude of fractionation to the masses of the isotopes. This question will be addressed elsewhere.

The relationship between the degree of evaporation and the

isotope mass fraction can be modelled using the Rayleigh equation: $R = f^{\alpha-1}$, where, for the case of ^{26}Mg ,

$$R = \frac{(^{26}\text{Mg}/^{24}\text{Mg})_{\text{sample}}}{(^{26}\text{Mg}/^{24}\text{Mg})_{\text{start}}}, \quad \alpha = \frac{(^{26}\text{Mg}/^{24}\text{Mg})_{\text{gas}}}{(^{26}\text{Mg}/^{24}\text{Mg})_{\text{solid}}}$$

and f is the fraction of the initial Mg that remains unevaporized. Here the fraction f is equivalent to the mass ratio of residual sample to starting material because Mg_2SiO_4 evaporates stoichiometrically. R and f are measured, then α is determined from the slope of a plot of $\ln f$ against $\ln R$, using the constraint that the regression line passes through the origin (R must be unity at $f = 1$). Table 2 shows four α values for each pair of isotopes, calculated from the data for 1,900, 1,940, 2,050 $^{\circ}\text{C}$ and for all temperatures combined. Figure 1 shows the isotopic compositions as a function of percent evaporation. Also shown are the Rayleigh curves calculated from the α values at 1,900 $^{\circ}\text{C}$. Comparison of the data at 1,900 and 2,050 $^{\circ}\text{C}$ (Tables 1 and 2) shows that larger isotope mass fractionations occur at 1,900 $^{\circ}\text{C}$ for Mg and Si, whereas the fractionation of O does not seem to depend on temperature. The data points at 1,940 $^{\circ}\text{C}$ do not lie on the curve defined by the data at 1,900 and 2,050 $^{\circ}\text{C}$, possibly owing to experimental error in this particular run (a portion of the liquid droplet may have fallen off during the experiment).

Hashimoto⁷ has studied the evaporation kinetics of Mg_2SiO_4 and has determined that the evaporation reaction is $\text{Mg}_2\text{SiO}_4(\text{s}, 1) \rightarrow 2\text{Mg}(\text{g}) + \text{SiO}_2(\text{g}) + [\text{O}_2(\text{g}) \text{ or } 2\text{O}(\text{g})]$. When isotope fractionation during evaporation is controlled by a

FIG. 2 Isotope mass fractionation of Mg and Si in CAIs^{1,2} and in liquid- Mg_2SiO_4 evaporation residues. F_{Mg} (‰ a.m.u.⁻¹) of CAIs is $\delta^{25}\text{Mg}$ or $\delta^{26}\text{Mg}/2$ (when the former value is not available), and F_{Si} is $\delta^{29}\text{Si}$ or $\delta^{30}\text{Si}/2$. F_{Mg} and F_{Si} of the experimental residues are $\delta^{25}\text{Mg}$ and $\delta^{29}\text{Si}$, relative to the starting material. The line is a linear regression of the Mg_2SiO_4 -residue data points.

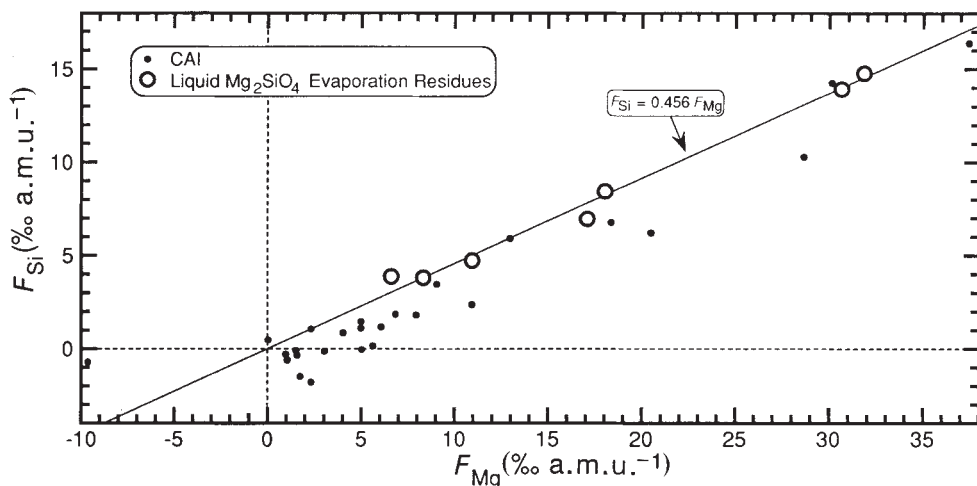


TABLE 2 Calculated gas/liquid partition coefficients

T (°C)	$^{18}\text{O}/^{16}\text{O}$	$^{17}\text{O}/^{16}\text{O}$	$^{26}\text{Mg}/^{24}\text{Mg}$	$^{25}\text{Mg}/^{24}\text{Mg}$	$^{30}\text{Si}/^{28}\text{Si}$	$^{29}\text{Si}/^{28}\text{Si}$
1,900	0.98047 ± 0.00029	0.98999 ± 0.00014	0.96786 ± 0.00051	0.98417 ± 0.00039	0.98512 ± 0.00022	0.99245 ± 0.00010
1,940	0.98497	0.99250	0.97230	0.98740	0.98801	0.99406
2,050	0.98033 ± 0.00052	0.99001 ± 0.00029	0.97078 ± 0.00013	0.98537 ± 0.00083	0.98650 ± 0.00063	0.99300 ± 0.00039
All temperatures	0.98063 ± 0.00050	0.99013 ± 0.00027	0.96962 ± 0.00082	0.98497 ± 0.00048	0.98599 ± 0.00042	0.99282 ± 0.00022
Theoretical	0.97694	0.98825	0.96077	0.97980	0.98374	0.99177

Uncertainties are 2σ from the uncertainty of the slopes of $\ln f$ plotted against $\ln R$.

kinetic isotope effect, the value of α is the inverse square root of the masses of the evaporating species. For the fractionation of ^{26}Mg relative to ^{24}Mg between gas and liquid (or solid) phases, $\alpha = \sqrt{24/26} = 0.9608$, because magnesium evaporates as the monatomic species. For $^{30}\text{Si}/^{28}\text{Si}$, $\alpha \approx \sqrt{60/62} = 0.9837$, because silicon evaporates as molecular SiO_2 , and $^{30}\text{Si}^{16}\text{O}_2$ and $^{28}\text{Si}^{16}\text{O}_2$ far outnumber other SiO_2 molecular species containing ^{30}Si and ^{28}Si , respectively. Finally, for $^{18}\text{O}/^{16}\text{O}$, $\alpha \approx (\sqrt{32/34} + \sqrt{60/62})/2 = 0.9769$, because oxygen evaporates as O_2 (here we assume diatomic rather than monatomic oxygen) and as SiO_2 molecules in equal numbers, and the numbers of $^{18}\text{O}^{16}\text{O} + ^{28}\text{Si}^{18}\text{O}^{16}\text{O}$ and $^{16}\text{O}_2 + ^{28}\text{Si}^{16}\text{O}_2$ are much larger than those of other O_2 and SiO_2 species containing ^{18}O and ^{16}O . These α values were used with the Rayleigh equation to construct the theoretical curves in Fig. 1. The experimental data for O, Mg and Si lie fairly close to the theoretical curves, but they give α values closer to unity than the theoretical values. There are two possible reasons for this discrepancy: (1) there are gas-phase species heavier than those inferred by Hashimoto⁷ involved in the evaporation reaction, such as MgO(g) ; (2) the residue is not isotopically homogeneous. The Rayleigh equation assumes homogeneity.

The slow diffusion rates of atoms in solids are almost certainly responsible for the lack of isotope mass fractionation during the evaporation of solid forsterite. Diffusion rates in liquid Mg_2SiO_4 are much larger, but may not be large enough to maintain complete homogeneity in the residue. To test this, the diffusion equation with surface evaporation as a boundary condition has been solved using the evaporation rates given by Hashimoto⁷, and diffusion coefficients of the elements in molten Mg_2SiO_4 calculated from the literature values of viscosity⁸ and electrical conductivity⁹. The result indicates that slower diffusion increases the effective α values, and the effect increases with temperature; among the three elements considered, O shows the least change in α with increasing temperature. Thus, to first order, the difference in isotope mass fractionation observed between 1,900 and 2,050 °C can be explained by the effect of diffusion, but a discrepancy between experiment and theory still remains after correction for this effect. This may point to the involvement of heavier gas species in the evaporation mechanism.

Our results have important implications for CAI formation in the primordial solar nebula. Figure 2 shows the relationship between isotope mass fractionation of Mg and Si in coarse-grained CAIs (literature values) and in Mg_2SiO_4 residues from our experiments. The CAI data lie close to the regression line fitted through the Mg_2SiO_4 data, but the CAIs generally have smaller F_{Si} or larger F_{Mg} values than the Mg_2SiO_4 residues. If coarse-grained CAIs were produced by partial evaporation of a precursor material that had the chondritic proportions of elements, they lost nearly equal fractions of the Mg and Si atoms contained initially¹⁰. This is a situation similar to that in our evaporation experiments (where equal fractions of the Mg, Si and O in Mg_2SiO_4 evaporated), although the chemical composition of the CAI precursor is different from that of forsterite. The agreement of isotope mass fractionation effects in the experiment with those in natural CAIs suggests that coarse-grained CAIs were produced by a non-equilibrium evaporation process

from molten precursor components containing abundant Mg and Si. The tendency of CAIs to have smaller F_{Si} or larger F_{Mg} than the experimental products is curious, as there is no experimental evidence for significant Mg evaporation before Si evaporation from chondritic materials¹⁰. Perhaps Si (or Mg) in the precursor material was slightly enriched in lighter (or heavier) isotopes relative to normal values.

It has been pointed out that FUN CAIs display much larger isotope mass fractionations of Mg, Si and O than ordinary coarse-grained CAIs, and yet the similarity of chemical composition between these two groups implies that their precursors experienced nearly equivalent degrees of evaporation. The FUN CAIs must have evaporated under non-equilibrium conditions, where the evaporation rate greatly exceeded the condensation rate, and at a temperature high enough for partial melting. Our experiments show that large mass fractionations and large degrees of evaporation can be produced in only a few minutes when the residue is molten. The ordinary CAIs may have been produced by evaporation of precursors at temperatures below the solidus.

The degree of fractionation of Mg and Si in the most highly fractionated FUN CAI, B7H10 of Brigham *et al.*², can only have been produced in a molten residue by evaporation of ~90% of the amounts of Mg and Si originally present (calculated from the fractionation factors in Table 2). This is also the amount of evaporation needed to change the chemical composition of the residue from that of a chondritic starting material to the CAI composition¹⁰. The isotopic composition of the first vapour generated by the earliest evaporation of Mg_2SiO_4 is predicted on the basis of the α values obtained to be $\delta^{25}\text{Mg} = -15.0\%$, $\delta^{17}\text{O} = -9.9\%$ and $\delta^{29}\text{Si} = -7.2\%$. (The lightest Mg isotopic composition of condensates collected by Esat *et al.*⁵ was $\delta^{25}\text{Mg} = -14.0\%$.) Fine-grained CAIs are enriched in the light isotopes of Mg (ref. 11), but by less than our first vapour. This is consistent with the formation of fine-grained CAIs by condensation from a vapour generated by evaporation of precursor materials having Mg with normal isotopic composition.

Although our experimental values of isotopic fractionation of Mg and Si closely match the observations in FUN and non-FUN CAIs, the O isotopic compositions in CAIs do not follow the mass fractionation trends seen in the experimental evaporations. The difference in behaviour between oxygen and the other major elements has been attributed to the presence of a large gaseous reservoir of oxygen compounds (such as CO , H_2O) in the solar nebula, which led to partial isotope exchange between gas and condensed phase³. If this exchange is corrected for, the ratio of O fractionation to Mg fractionation (compared to normal CAIs) is similar in most FUN inclusions and in the Mg_2SiO_4 residues. This implies that evaporation must have taken place either before or during the existence of the gaseous nebula, but not after its dissipation. □

Received 5 March; accepted 22 August 1990.

1. Clayton, R. N., Hinton, R. W. & Davis, A. M. *Phil. Trans. R. Soc. A* **325**, 483–501 (1988).
2. Brigham, C. A., Hutcheon, I. D., Papanastassiou, D. A. & Wasserburg, G. J. *Proc. lunar planet. Sci.* **19**, 132–133 (1988).
3. Clayton, R. N. & Mayeda, T. K. *Geophys. Res. Lett.* **4**, 295–298 (1977).
4. Molini-Velsko, C., Mayeda, T. K. & Clayton, R. N. *Proc. lunar planet. Sci.* **18**, 657–658 (1987).
5. Esat, T. M., Spear, R. H. & Taylor, S. R. *Nature* **319**, 576–578 (1986).

6. Somorjai, G. A. & Lester, J. E. *Prog. Solid St. Chem.* **4**, 1-53 (1967).
7. Hashimoto, A. *Nature* **347**, 53-55 (1990).
8. Bockris, J. O'M., Mackenzie, J. D. & Kitchener, J. A. *Trans. Faraday Soc.* **51**, 1734-1748 (1955).
9. Bockris, J. O'M., Kitchener, J. A., Ignatowicz, S. & Tomlinson, J. W. *Trans. Faraday Soc.* **48**, 75-91 (1952).
10. Hashimoto, A. *Geochem. J.* **17**, 111-145 (1983).
11. Brigham, C. A., Papanastassiou, D. A. & Wasserburg, G. J. *Proc. lunar planet. Sci.* **16**, 93-94 (1985).

ACKNOWLEDGEMENTS. We thank J. A. Wood for discussions. This work was supported by NASA and the NSF.

Photoelectrochemical information storage using an azobenzene derivative

Z. F. Liu, K. Hashimoto & A. Fujishima

Department of Synthetic Chemistry, Faculty of Engineering,
University of Tokyo, Hongo, Tokyo 113, Japan

HIGH-DENSITY information storage is becoming an increasingly important technological objective. The 'heat-mode' storage techniques (in which only the thermal energy of laser light is used in the recording process and hence information usually stored as a physical change of the storage media) that are used in current optical memories are limited by the diffraction properties of light¹, and the alternative 'photon-mode' (in which information is stored as a photon-induced chemical change of the storage media) has attracted attention recently for high-density storage. The most promising candidates for realizing this mode seem to be photochromism and photochemical hole burning; but these have some intrinsic drawbacks^{1,2}. Here we present a novel 'photon-mode' technique that uses the photoelectrochemical properties of a Langmuir-Blodgett film of an azobenzene derivative. The system can be interconverted photochemically or electrochemically between three chemical states, and this three-state system is shown to provide a potential storage process that allows for ultra-high storage density, multi-function memory and non-destructive information readout.

Two types of reversible processes, photochemical *cis-trans* isomerization (see, for example, ref. 3) and electrochemical oxidation-reduction⁴, are available in an azobenzene system.

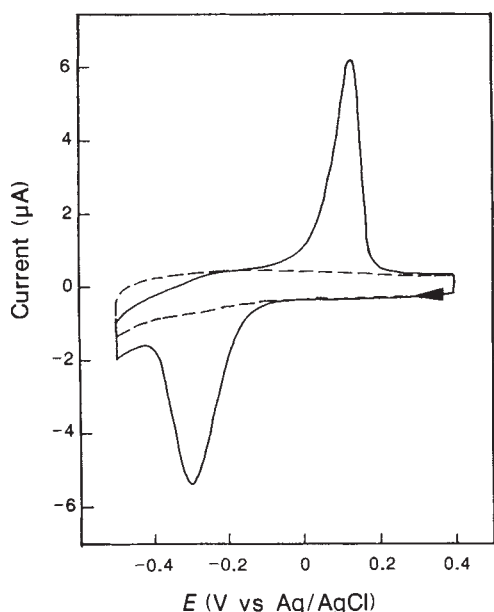


FIG. 1 Cyclic voltammograms of *trans* (dashed line) and *cis* (solid line) ABD monolayer films on a SnO_2 glass electrode. The sweep rate was 20 mV s^{-1} .

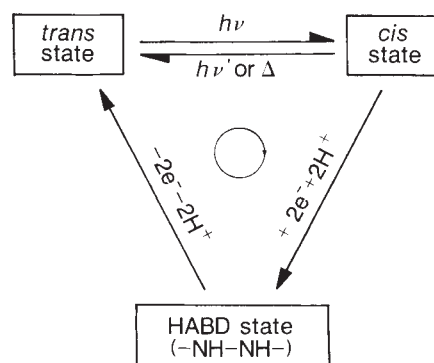


FIG. 2 Reaction mechanism of ABD molecules.

By using both processes, it is possible to store information photoelectrochemically. The proposed storage system is based on two observations: (1) the *cis* azobenzene derivative used can be reduced to a hydrazobenzene species at substantially more anodic potentials than the *trans* form; (2) the hydrazobenzene derivative is exclusively oxidized to the *trans* form.

The azobenzene derivative (ABD) used was 4-octyl-4'-(5-carboxy-pentamethylene-oxy)-azobenzene. It was deposited in the *trans* form as a monolayer film onto a transparent SnO_2 glass substrate, the working electrode, using the Langmuir-Blodgett method. The surface pressure and sub-phase (0.2 mM CdCl_2 aqueous solution) temperature during film fabrication were controlled at 25 mN m^{-1} and 20°C , respectively. We studied the electrochemical behaviour using cyclic voltammetry in a 0.2 M aqueous potassium perchlorate, buffered to pH 7.0 with Britton-Robinson solution. The reference electrode was Ag/AgCl and a Pt wire was used as the counter electrode. A 500-W xenon lamp and a 14-mW He-Cd laser (325 nm) were used to induce general and localized isomerization, respectively, of ABD molecules. The ultraviolet output (320-380 nm) of the lamp was separated using a glass filter.

Figure 1 compares the cyclic voltammetric behaviour of *trans*- and *cis*-ABD monolayer films. (The *cis*-state film was obtained by irradiating the *trans*-state film with the xenon lamp for one minute.) In the applied potential range no reduction and oxidation peaks were observed in the *trans*-state film, whereas in the *cis*-state film significant peaks were obtained from the reduction of *cis*-state film to the hydrazobenzene derivative (HABD, $-\text{NH}-\text{NH}-$) followed by the oxidation of the HABD-state film⁵. A similar, but slow reduction of the *trans*-state film also occurred at more negative potential, as evidenced by the broad reduction peak and the following oxidation peak (not shown in Fig. 1). The difference in reduction potential ($E_{1/2}$) of the *trans*- and *cis*-state films was estimated to be $>400 \text{ mV}$ in neutral solution. On the other hand, the HABD-state film was exclusively oxidized to the thermodynamically stable *trans*-state film because no electrochemical activity was observed after its oxidation in the electro-inactive potential range of *trans*-state film. Similar photoelectrochemical behavior was also obtained for some other long-chain fatty acid derivatives of azobenzene that have chromophores located at different places on their alkyl chains. Our result differs from those for azobenzene itself⁴, for which *trans* and *cis* molecules are reduced at the same potential in aqueous solution. This difference might be the result of structural or orientational differences of the surface-bound azobenzenes.

Figure 2 shows the three-state, one-way cyclic behaviour of the photochemical and electrochemical processes. This photoelectrochemical hybrid process is evidence that a 'photoelectrochemical memory' can be created. The thermally stable HABD-state can be used as the 'storage state' for information

instead of the thermally unstable *cis*-state film. We have made preliminary investigations of the stability of the 'storage-state' molecules under several conditions. Although HABD was slowly oxidized to *trans*-ABD in the presence of oxygen, no change was observed on the experimental timescale (~ 5 h) under de-aerated or negatively (-0.3 V) biased conditions. The formation of phenyl hydrazines or anilines by further reduction of hydrazobenzenes⁶ was not observed, probably because of the relatively small cathode potential used.

The photochemical-electrochemical hybrid feature of our system gives rise to two possible means of high-density storage, high-resolution 'optical writing' and high-resolution 'electrochemical writing'. For optical writing, high-density storage can be obtained using a laser with a narrow beam to induce a local *trans*-*cis* isomerization, followed by switching the potential of the entire film to a voltage where only *cis*-ABD is reduced. Figure 3 shows the cyclic voltammograms when a He-Cd laser (beam diameter of ~ 1 mm) was used to irradiate six different areas. The number of electrochemically reacting ABD molecules was calculated by integrating the anode current/potential curves, which showed a linear increase proportional to the increase in the number of irradiated areas (Fig. 4). The small change in the average number of reacting molecules per area is attributed to the thermal *cis*-*trans* isomerization in the previously irradiated areas. A separate experiment was performed to determine the stability of the *cis*-state film in the dark (Fig. 4), and it is obvious that only illuminated areas show an electrochemical response. Consequently, the storage density depends only on the irradiated area, and if such an operation were carried out at the diffraction limit of laser light, one would be able to store 10^8 bits cm^{-2} in a two-dimensional configuration. Moreover, if several azobenzene species having similar hybrid behaviour but substantially different spectral responses were located in the storage film, the storage density could be further increased by using a tunable laser.

High-density electrochemical writing, on the other hand, is

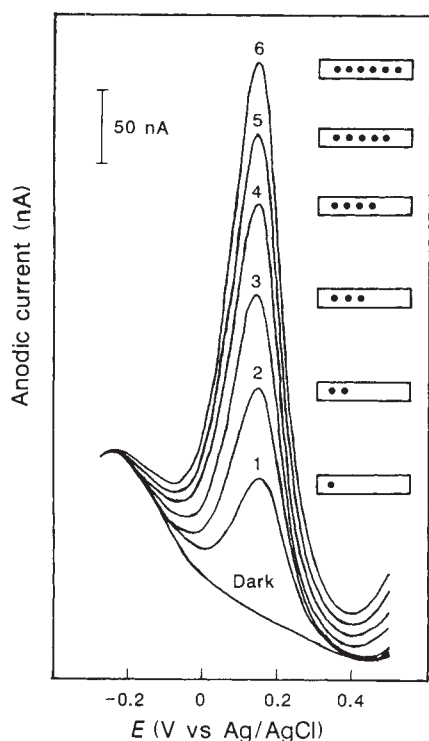


FIG. 3 Anodic peaks of the cyclic voltammograms for different areas of irradiation. The symbol ● represents the unit area irradiated for 10 s.

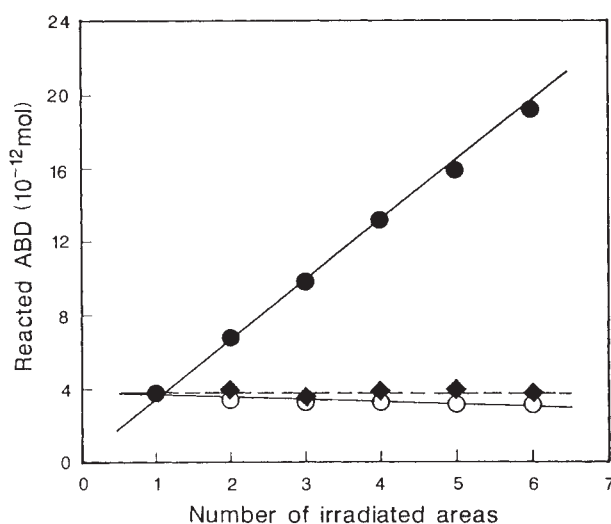


FIG. 4 Relationship between the reacting ABD and the number of irradiated areas. ●, Total number of reacting ABD molecules; ○, average number per irradiated area; ◆, corrected average number per irradiated area, in which the contribution of the thermal *cis*-*trans* isomerization has been eliminated.

obtained by controlling the reduction area of *cis*-state film after uniform irradiation of the entire film. This is possible using a sharp glass-coated probe as the counter electrode, with the distance between the probe and the storage-film-modified electrode being controlled. The probe can be positioned using the same technique as in scanning tunnelling microscopy⁷. In principle, a storage density of 10^{12} bit cm^{-2} is possible in a two-dimensional configuration, without using the frequency domain of light⁸.

Using our two-step photoelectrochemical hybrid process also offers a realistic approach to non-destructive reading of the information stored. There is a large difference in the absorption spectra of *trans*- and HABD-state films owing to the conversion of $-\text{N}=\text{N}-$ to $-\text{NH}-\text{NH}-$ in the ABD chromophore. Thus the information stored in either writing mode can be read by monitoring this spectral change. Because both photochemical isomerization and electrochemical reduction are necessary for formation of the HABD storage state, optical reading will not destroy the stored information. Such a hybrid process has a significant advantage over the existing 'photon-mode' techniques, which are based on a one-photon process and are destructive.

The three-state feature of this cyclic process may allow a multi-function memory to be realized because the metastable *cis*-state film can be used for short-term storage, and the HABD-state film for long-term storage. Finally, the stored information can be erased by anodizing the entire film, whereupon the hydrazobenzene derivative is re-oxidized to the original *trans*-form, leading to a construction of a reversible memory device. The reversible durability has been studied by repeating an isomerization $\rightarrow$ reduction $\rightarrow$ oxidation operation. No discernible change in the nature of the reaction was observed during several hundred cycles, although a small decrease in the oxidation-reduction peak currents occurred owing to a dissolution of the film molecules in the aqueous electrolyte. This dissolution problem might be overcome in future by using a solid electrolyte rather than an aqueous one.

Our idea opens several new avenues for information storage because the $-\text{C}=\text{N}-$ and $-\text{C}=\text{C}-$ compounds have similar photochemical and electrochemical behaviour. In future studies design of the molecules and of the matrix will play an important

part in obtaining a system having a long-wavelength response so that available semiconductor lasers can be used. □

Received 2 July; accepted 29 August 1990.

1. Gutierrez, A. R., Friedrich, J., Haarer, D. & Wolfum, H. *IBM J. Res. Dev.* **26**, 198–208 (1982).
2. Ando, E., Miyazaki, J. & Morimoto, K. *Thin Solid Films* **133**, 21–28 (1985).
3. Griffiths, J. *Chem. Soc. Rev.* **1**, 481–493 (1972).
4. Laviron, E. & Mugnier, Y. *J. electroanal. Chem.* **111**, 337–344 (1980).
5. Liu, Z. F., Hashimoto, K. & Fujishima, A. *J. electroanal. Chem.* (in the press).
6. Holleck, L., Vavricka, S. & Heyrovsky, M. *Electrochim. Acta* **15**, 645–656 (1970).
7. Binnig, G. & Rohrer, H. *Rev. mod. Phys.* **59**, 615–625 (1987).
8. Kwak, J. & Bard, A. J. *Anal. Chem.* **61**, 1794–1799 (1989).

Warming trend in the western Mediterranean deep water

J. P. Bethoux, B. Gentili, J. Raunet & D. Tailliez

Laboratoire de Physique et Chimie Marines, Université de Paris 6, UA CNRS, BP 8, F-06230 Villefranche sur Mer, France

THE western Mediterranean Sea comprises three water masses: a surface layer (from 0 to ~150 m depth), an intermediate layer (~150–400 m) issuing from the eastern basin, and a deep water mass at depths below 400 m. The deep water is homogeneous and has maintained a more or less constant temperature and salinity from the start of the century until recently¹. Here we report measurements from the Medatlante cruises of December 1988 and August 1989, which show the deep layer to be 0.12 °C warmer and ~0.03 p.s.u. more saline than in 1959. Taking these data together with those from earlier cruises, we find a trend of continuously increasing temperatures over the past three decades. These deep-water records reflect the averaged evolution of climate conditions at the surface during the winter, when the deep water is formed. Consideration of the heat budget and water flux in the Mediterranean^{2,3} leads to the possibility that the deep-water temperature trend may be the result of greenhouse-gas-induced local warming.

During the 1960s and 1970s, the formation of deep water in the northwestern Mediterranean was studied^{4,5}. This process occurs in late winter after a preconditioning phase in which intense cooling and evaporation increase the surface density^{4,5}. Deep convective mixing of surface and intermediate layers then create the deep water. Analysis of extensive hydrographic data¹ obtained between 1959 and 1982 has shown interannual variations in the characteristics of the deep water at depths $\geq 2,000$ m, but there has also been an increase in the mean potential temperature (Fig. 1). Further results from Phycmed (1981) cruise, Mediprod 5 (1985) and Medatlante cruises (JGOFS-France programme) have confirmed this increase

throughout the deep water of the western basin. Between 1959 and 1989 the mean annual increase of the potential temperature at depths $\geq 2,000$ m, dT_d , was 4×10^{-3} °C yr⁻¹. Salinity in the deep layer also increased by ~0.03 p.s.u. and the density therefore remained constant at ~1,029.10 kg m⁻³. In the western basin (area 8.5×10^{11} m²), the temperature increase, assuming a volume of deep water of $\sim 11 \times 10^{14}$ m³ (for depth > 400 m), the above density and a specific heat of 3.97×10^3 J kg⁻¹ °C⁻¹, corresponds to a mean change in heat content $Q_{Td} = 0.67$ W m⁻².

Changes in the temperature and salinity in various parts of the world ocean have been reported^{6–9} but there are many coupled causes and feedbacks through atmospheric circulation, marine mesoscale turbulence and advection from adjacent or distant basins. In the land-locked Mediterranean Sea, the characteristics of the deep western water result from air-sea exchanges and dense water formation (Fig. 2) in the north-western area. In the eastern basin, formation of dense water in winter also occurs, transforming surface water into homogeneous deep water that fills the basin to ~400 m depth and then flows out of the basin over the Sicily-Tunisia sills, becoming western intermediate water. From an average over many years, the annual deep outflow ($V_d = 50 \times 10^{12}$ m³ yr⁻¹ at mean temperature T_d) over the Gibraltar sill is compensated for by the winter transformation of surface water ($V_s = 12 \times 10^{12}$ m³ yr⁻¹ at temperature T_s) and by the inflow of intermediate water ($V_i = 38 \times 10^{12}$ m³ yr⁻¹ at mean temperature T_i)³. Consequently, the temperature change of deep and intermediate water, dT_d and dT_i , may be related to winter surface temperatures T_{sw} and T_{se} in the western and eastern basins, respectively, and to intermediate temperature T_i by

$$\text{Vol}_w \frac{dT_d}{dt} = V_s T_{sw} + V_i T_i - (V_s + V_i) T_d$$

$$\text{Vol}_e \frac{dT_i}{dt} = V_i (T_{se} - T_i)$$

where Vol_w (11×10^{14} m³) and Vol_e (18×10^{14} m³) represent the volumes of deep water and T_{sw} and T_{se} the surface temperatures in the western and eastern basins, respectively. The value of dT_d may be related to the linear change in the surface temperature which is assumed equal in the two basins, $T_{sw} = T_{sw0} + \alpha t$, and $T_{se} = T_{se0} + \alpha t$, where T_{sw0} and T_{se0} are the initial surface temperatures in the two basins. The mean value of dT_i and the α coefficient are determined from the solution of the two differential equations, this solution being mean-square-root adjusted to the measured deep temperatures in Fig. 1. From the proposed volumes and flux, the western deep-water warming ($dT_d = 4 \times 10^{-3}$ °C yr⁻¹) is linked to surface temperature increase dT_s of 20×10^{-3} °C yr⁻¹ and to an intermediate-layer warming dT_i of $\sim 5 \times 10^{-3}$ °C yr⁻¹. Seasonal variability of surface temperature is much higher than the calculated dT_s and a tem-

FIG. 1 Potential temperature plotted against time for the bottom water at depths $\geq 2,000$ m in the western Mediterranean. Open square, mean potential temperature and error bars; rhomb, minimum potential temperature with error bars. Cruises and data from 1959 to 1982 are listed in ref. 1; Phycmed (1981) data from R. Chesselet and J. P. Gouillou (personal communication); Mediprod 5 (1985) data from L. Prieur (personal communication).

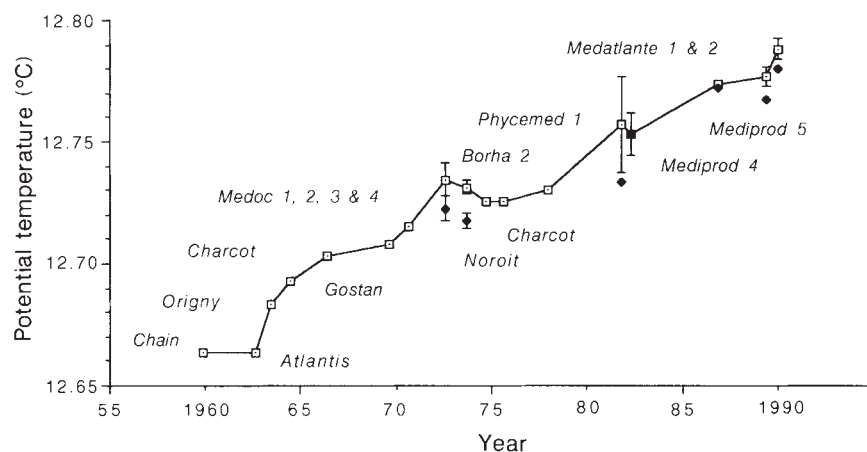
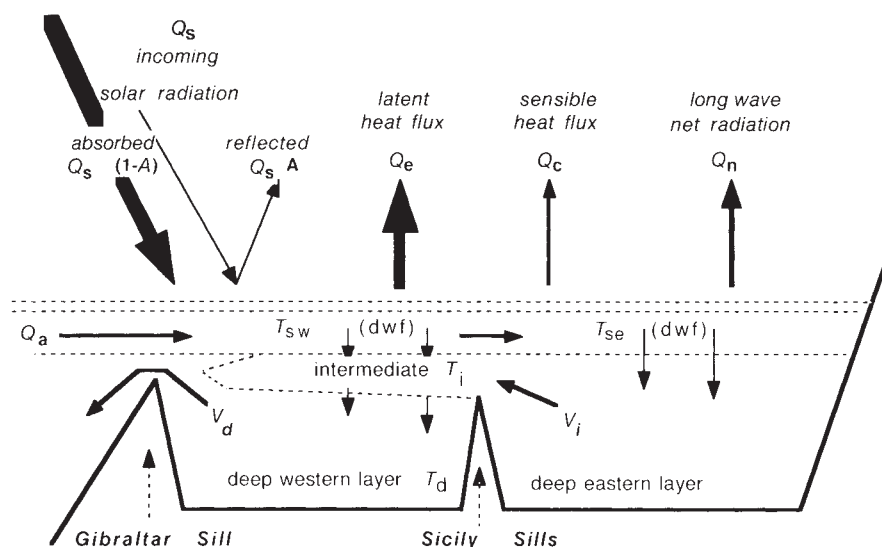


FIG. 2 Air-sea heat exchanges across the Mediterranean surface (upper part) and resulting heat and water flux in the western and eastern basins (lower part). The downward arrows indicate the dense-water formation (dwf), which carries heat and water from the surface to the deep layers.



perature change in the intermediate layer is not yet apparent, in either the eastern basin, where there are few historical data, or in the western basin, where continuous mixing with surface water induces spatial and temporal variabilities which may hide a small thermal trend. But if T_i and T_{se} are assumed constant, that is, if the thermal change originates only in the western surface layer, the surface temperature increase since 1959 would be 2.5°C , an unrealistic value. This result strengthens the hypothesis of a temperature increase of the whole Mediterranean surface layer during the winter. In the western basin, the temperature change of the flux of the intermediate water corresponds to a change in the heat content Q_{Ti} of 0.03 W m^{-2} . In the surface layer ($1 \times 10^{14}\text{ m}^3$), the dT_{sw} corresponds to a $Q_{T_{sw}}$ of 0.30 W m^{-2} . The total change in the heat content of the entire water column deduced from the temperature change of the deep water and the fluxes is $Q_{tot} = Q_{Td} + Q_{Ti} + Q_{T_{sw}} = 1.0\text{ W m}^{-2}$. According to the hypotheses, the change in the heat content of the eastern basin is the same.

The heat budget of the entire Mediterranean may be written as $Q_{tot} = Q_s(1-A) + Q_a - Q_c - Q_e - Q_n$ where Q_s is the solar heat input, A is the mean albedo, Q_a is the heat gained by marine advection at the Straits of Gibraltar, Q_c is the sensible (conduction) heat, Q_e the latent (evaporation) heat and Q_n the net infrared radiation. The advection term Q_a , the thermal excess of inflowing Atlantic waters over the deep Mediterranean outflow, is relatively small, 6.5 W m^{-2} , compared with the solar radiation, 195 W m^{-2} (refs 2, 3). In thermal equilibrium, characterized by a constant mean annual temperature, $Q_{tot} = 0$, the heat gains $Q_s(1-A) + Q_a$ are entirely transferred to the atmosphere by net back radiation (68.5 W m^{-2}) and losses of sensible (13.5 W m^{-2}) and latent (119.5 W m^{-2}) heat. This latter corresponds to an annual loss of evaporated water of 1.54 m , which is greater than the estimated freshwater gain of $\sim 0.60\text{ m}$. This leads to concentration basin dynamics, characterized by the formation of dense water in winter, superposed and reversed flows over the Sicily and Gibraltar sills and by homogeneous deep waters in the eastern and western basins.

The constant water densities at intermediate depth in the Atlantic and in the Mediterranean, on each side of the Straits of Gibraltar, ensure steady-state water dynamics. Moreover, climate data from the Mediterranean have not shown a persistent increase in solar radiation. Q_{tot} and the surface warming are therefore linked to an increase of marine advection and/or to

a decrease of heat transfers to the atmosphere. From the constant water flux through the Straits of Gibraltar³ and the area of the Mediterranean ($2.5 \times 10^{12}\text{ m}^2$), a possible change in the marine advection may be written as $dQ_a = 2.7(dT_{Atl} - 0.95 dT_d)\text{ W m}^{-2}$, where dT_{Atl} ($^\circ\text{C yr}^{-1}$) is the temperature change of inflowing Atlantic waters and dT_d is the previously determined temperature change of deep water in the western basin. An advection increase of 1 W m^{-2} (the calculated change in heat content of water column) would need a warming of the Atlantic surface water of $0.37^\circ\text{C yr}^{-1}$, an unrealistic hypothesis. A dT_{Atl} of the order of dT_s or of the global surface air temperature change dT_a ($\sim 6 \times 10^{-3}\text{ }^\circ\text{C yr}^{-1}$ during the past thirty years, ref. 10) corresponds to a negligible dQ_a when compared with Q_{tot} . Consequently, the prime cause of the Q_{tot} evolution concerns the air-sea exchanges.

Climate records have not shown continuous evolution of wind speed, relative moisture and cloudiness during the past fifty years, a period covering the residence time of deep western water and the deep warming record. When these meteorological parameters are assumed constant, the air-sea exchange evolutions are only linked to the changes of air and sea temperatures and radiatively active gas concentrations in the atmosphere. In a previous study², atmospheric and marine data acquired from a buoy anchored in the Algero-Provençal Basin from 1964 to 1970 allowed the seasonal air-sea heat transfers to be calculated. The drag coefficient used in the bulk aerodynamic method was constrained to give latent and sensible heat fluxes balancing the sum of net back radiation, solar radiation and advection, the two latter terms being well known in the Mediterranean area. First approximations of the small mean changes in Q_c and Q_e driven by small air and sea temperature changes, dT_a and dT_s , are $dQ_c = 16.8(dT_s - dT_a)$ and $dQ_e = 32.5(dT_s - 0.77 dT_a)$ where dQ_c and dQ_e are in W m^{-2} , and dT_s and dT_a are in $^\circ\text{C yr}^{-1}$.

Similarly, the change in the net long-wave radiation calculated from a Laevastu¹¹ type formula, where the cloudiness and relative moisture are assumed constant, is approximated by a function of T_s derived from the heat budget study and a function describing the change of the radiatively active gas concentration in the atmosphere $dQ_n = -0.60 dT_s - x dC$ where dQ_n is in W m^{-2} , dT_s is in $^\circ\text{C yr}^{-1}$, and the last term, $x dC$, constitutes the greenhouse effect in W m^{-2} . The three last relations show positive or negative feedbacks between air and sea temperatures and heat transfers. The simplified heat-budget evolution in the

Mediterranean Sea, neglecting solar and/or advection evolution, is $Q_{\text{tot}} = -dQ_c - dQ_e - dQ_n$ or, expressed as a function of dT_s and dT_a (with $Q_{T_{\text{sw}}} = 15 dT_s$), $x dC = 0.7 + 63.7 dT_s - 41.8 dT_a$.

From the previous estimates of dT_s ($20 \times 10^{-3} \text{ }^\circ\text{C yr}^{-1}$), the greenhouse effect is zero if the air warming (driven by processes outside of the basin) is $\sim 5 \times 10^{-2} \text{ }^\circ\text{C yr}^{-1}$, eight times the mean global change¹⁰. Inversely, the maximum calculated greenhouse effect (for $dT_a = 0$) is 2 W m^{-2} . But, in the Mediterranean Sea, an area almost without external thermal advection, sea surface and air temperatures are likely to change similarly, as previously shown at a global scale¹². An upper estimate of dT_a is dT_s and a lower estimate is the global evolution of T_a (ref. 10). These temperatures determine a greenhouse effect over the Mediterranean Sea of $1.1\text{--}1.7 \text{ W m}^{-2}$.

These effects represent an evolution of 2% in Q_n , which acts to warm the sea and to a lesser extent to increase the sensible and latent transfers. Consequently, for a dT_a of $6 \times 10^{-3} \text{ }^\circ\text{C yr}^{-1}$, salinity increases by 2×10^{-3} p.s.u. yr^{-1} in the surface layer. A temperature-type calculation links the salinity change in the deep layer to a surface salinity increase of $\sim 5 \times 10^{-3}$ p.s.u. yr^{-1} , which requires, besides the evaporation effect, a reduction of mean freshwater input by $\sim 2\%$ for a further salinity effect of 3×10^{-3} p.s.u. yr^{-1} . On the scale of the whole sea, the high spatial and temporal variability of precipitation prevents any direct evidence of a general decrease, but analysis of data acquired at Monaco between years 1911 and 1985¹³ shows decreasing precipitation which, during the past 30 years, reduced the mean freshwater input by 7%. The marine estimated greenhouse effect is slightly lower than the global value 1.9 W m^{-2} calculated from the evolution of the radiative-gas concentration in the atmosphere since 1860¹⁴. The dynamical coupling of the two large eastern and western basins, where the deep waters have different residence times (about 20 years in the western basin and 50 years in the eastern basin), may somewhat reduce the deep thermal evolution when the driving force is not steady. For instance, the global change dT_a (ref. 10) shows a stagnation between 1940 and 1965, which, if it applies to the Mediterranean area, may result in a constant temperature in the intermediate layer during the past decade. Moreover, some new flow values are smaller than ours (annual Atlantic surface inflow of $41 \times 10^{12} \text{ m}^3 \text{ yr}^{-1}$ (ref. 15), and deep outflow of $27 \times 10^{12} \text{ m}^3 \text{ yr}^{-1}$ for a mean outflow salinity of 37.9 p.s.u. (ref. 16), instead of our values 53 and $50 \times 10^{12} \text{ m}^3 \text{ yr}^{-1}$, respectively); these would lead to an increase in the surface warming and the calculated greenhouse effect would reach maximum values (for $dT_a = 0$) of 2.3 and 2.6 W m^{-2} respectively, not very different from our value of 2 W m^{-2} . Nevertheless, the Mediterranean Sea offers the opportunity to quantify the climate changes and particularly the greenhouse effect through the temperature change in the western deep water. This homogeneous deep water constitutes a rather stable integrating medium and oceanographic measurements may complement the atmospheric experiments, which are complicated by high spatio-temporal variabilities of climate. $\square$

Derivation of some modern arc magmas by melting of young subducted lithosphere

Marc J. Defant* & Mark S. Drummond†

* Department of Geology, University of South Florida, Tampa, Florida 33620, USA

† Department of Geology, University of Alabama, Birmingham, Alabama 35294, USA

MOST volcanic rocks in modern island and continental arcs are probably derived from melting of the mantle wedge, induced by hydrous fluids released during dehydration reactions in the subducted lithosphere¹. Arc tholeiitic and calc-alkaline basaltic magmas are produced by partial melting of the mantle, and then evolve by crystal fractionation (with or without assimilation and magma mixing) to more silicic magmas²—basalt, andesite, dacite and rhyolite suites. Although most arc magmas are generated by these petrogenetic processes, rocks with the geochemical characteristics of melts derived directly from the subducted lithosphere are present in some modern arcs where relatively young and hot lithosphere is being subducted. These andesites, dacites and sodic rhyolites (dacites seem to be the most common products) or their intrusive equivalents (tonalites and trondhjemites) are usually not associated with parental basaltic magmas³. Here we show that the trace-element geochemistry of these magmas (termed 'adakites') is consistent with a derivation by partial melting of the subducted slab, and in particular that subducting lithosphere younger than 25 Myr seems to be required for slab melting to occur.

Gill¹ and Drummond and Defant³ have discussed the expected geochemical characteristics of rocks derived from the melting of subducted basalt. Experimental work indicates (see ref. 3 for a review) that the partial melting of metamorphosed basalt (amphibolite and eclogite) will produce relatively high-Al, mostly corundum-normative silicic melts ($\text{Al}_2\text{O}_3 > 15\%$ at 70% SiO_2). The existence of garnet or amphibole as a residual will lead to melts with low concentrations of yttrium and heavy rare-earth elements (HREEs).

The age of the lithosphere^{4,5} subducted along trenches associated with 12 island arcs is plotted against the concentrations of Y and normalized Yb in volcanics sampled from these arcs in Fig. 1. In all but nine samples (>400 samples plotted), the Y and normalized Yb values of the island-arc andesites, dacites and sodic rhyolites (ADRs) are ≥ 20 . Island arcs near continental margins, where subducted crustal sediments have been documented (such as Lesser Antilles arc⁶, Luzon arc, Philippines⁷, and Aleutian arc⁸) have been omitted because of the potential of lower Y and Yb values by source contamination. In these cases, however, only a few samples have Y and normalized Yb values < 20 .

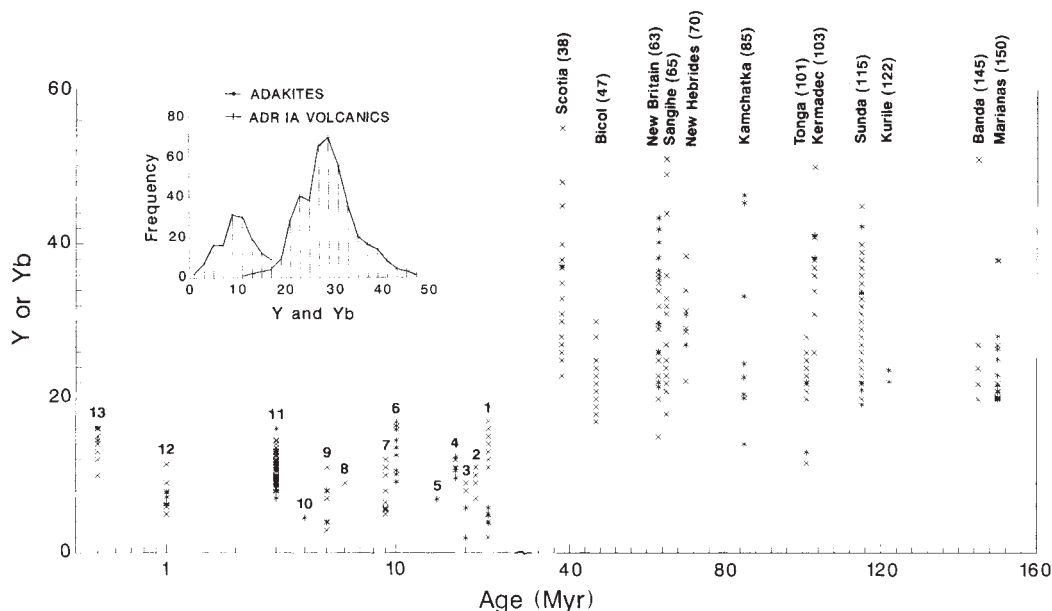
Thirteen Cenozoic arc ADRs associated with young subducted crust (the crust subducted at the trench is ≤ 25 Myr old) have also been plotted in Fig. 1. All have Y and normalized Yb values < 19 , and most are < 15 . There are other volcanics that may fall into this group (such as the New Guinea Highlands and Eastern Papua, Papua New Guinea⁹, Kadavu Island, Fiji (ref. 10 and J. B. Gill, personal communication), and the Abu Volcano Group, Japan¹¹) but either the age of the subducted lithosphere is unknown or data are unavailable for these regions. Two general observations can be made. First, it is rare to find low Y and Yb concentrations in island-arc ADRs associated with subducted lithosphere older than 25 Myr. Second, in arcs where young crust is being subducted, ADR samples with both high and low Y and Yb can be found. For example, Mount St Helens has mostly low Y and Yb values but associated volcanoes in the Cascade range (not shown on the diagram because the Cascade arc is not an island arc) have higher values, comparable

Received 17 April; accepted 22 August 1990.

1. Lacombe, H., Tchernia, P. & Gamberoni, L. *Prog. Oceanogr.* **14**, 319–338 (1985).
2. Bethoux, J. P. thesis Univ. Paris (1977).
3. Bethoux, J. P. *Oceanol. Acta* **3**, 79–88 (1980).
4. Medoc Group Nature **227**, 1037–1040 (1970).
5. Tchernia, P. *Colloques Int. CNRS* **215**, 17–21 (1974).
6. Cane, M. A. *Science* **222**, 1189–1195 (1983).
7. Brewer, P. G. et al. *Science* **222**, 1237–1239 (1983).
8. Roemmich, D. & Wunsch, C. *Nature* **307**, 447–450 (1984).
9. Lazier, J. R. N. *Deep Sea Res.* **35**, 1247–1253 (1988).
10. Jones, P. D. et al. *Nature* **332**, 790 (1988).
11. Laevastu, T. *Soc. Sci. Fennica* **25**, 1–136 (1960).
12. Folland, C. K., Parker, D. E. & Kates, F. E. *Nature* **310**, 670–673 (1984).
13. *Un siècle d'observations météorologiques à Monaco* Vol. 3 (Centre Scientifique de Monaco, Monaco, 1988).
14. Mitchell, J. F. B. *Rev. Geophys.* **27**, 115–139 (1989).
15. Perkins, H., Kinder, T. & La Violette, P. *J. phys. Oceanogr.* **20**, 242–263 (1990).
16. Bryden, H. L. & Pillsbury, R. D. *Computational Mech. Publs* (in the press).

ACKNOWLEDGEMENTS. This is a contribution to CNRS/INSU/JGOFs programme.

FIG. 1 Y(*) and Yb(x) content of arc volcanics and plutonics plotted against the estimated age of the subducting lithosphere. Yb has been chondrite normalized²⁷ and multiplied by 2.4 to make the values approximately equivalent to Y values. Ages for the subducting crust^{4,5} associated with the Quaternary island arcs are given after the names in parentheses. The adakites are represented by numbers: 1, El Valle and La Yeguada volcanoes, Panama (25 Myr)^{13,28}; 2, Cook Island, Chile (22 Myr)²⁹; 3, Paracale pluton, Philippines (20 Myr) (M. J. D., unpublished data; samples from U. Knittel and W. Winter); 4, Austral Andes (18 Myr)^{30,31}; 5, Aleutian arc, dredged sample, Alaska (15 Myr)¹²; 6, Skagway batholith, Alaska (10 Myr)³²; 7, North Kamchatka (9 Myr)³³; 8, Sierra Madre, Mexico (6 Myr)³⁴; 9, Patagonia, Chile (5 Myr) (ref. 35 and D. S. Bartholomew, personal communication); 10, Los Chocoyos, Guatemala (4 Myr)³⁶; 11, Mount St Helens (3 Myr)³⁷; 12, Baja of California, Mexico (1 Myr)³⁸; 13, Woodlark Basin (0.5 Myr)³⁹.



to those of the island arcs shown in Fig. 1.

We use the term adakite to describe these Cenozoic arc magmas because they were first documented in Adak Island (Aleutian Islands, Alaska)¹². They are volcanic or intrusive rocks in Cenozoic arcs associated with subduction of young (≤ 25 Myr) oceanic lithosphere. Adakites are characterized by $\geq 56\%$ SiO_2 , $\geq 15\%$ Al_2O_3 (rarely lower), usually $< 3\%$ MgO (rarely above 6% MgO), low Y and HREE relative to island-arc ADRs (for example, Y and Yb ≤ 18 and 1.9 p.p.m., respectively), high Sr relative to island-arc ADR (rarely < 400 p.p.m.), low high-field strength elements (HFSEs), as in most island-arc ADRs, and $^{87}\text{Sr}/^{86}\text{Sr}$ usually < 0.7040 . The petrography of adakites is somewhat variable. Plagioclase is a ubiquitous phase, and amphibole is common in all samples but the more MgO -rich varieties. Clinopyroxene and orthopyroxene are also common constituents, with biotite and opaques occurring frequently. A common assemblage is plagioclase and amphibole (with or without biotite, pyroxene and opaques). Adakites are rarely associated with parent basalts or basaltic andesites. When they are, the basaltic rocks are usually extremely enriched in large-ion lithophile elements (such as absarokites and shoshonites¹³).

Trace-element modelling of the partial melting of basalt transformed to eclogite (with or without amphibolite) can explain the low Yb and Y values of adakites in Fig. 1. Various curves derived from fresh and altered MORB have been superimposed on a plot of Sr/Y against Y in Fig. 2. Most recent (Quaternary) island-arc ADRs do not resemble adakites even though both groups have comparable silica levels (Fig. 2). These arc volcanics seem to be derived from basaltic magmas originating from a metasomatized mantle-wedge source, perhaps enriched with a small sediment component. Basaltic melts, which have Sr and Y values comparable to adakites, probably differentiate by assimilation and fractional crystallization (AFC) (and magma mixing) towards more silicic compositions^{2,14}. The partial melting of a periodotitic source and subsequent differentiation is toward high Y concentrations (and low Sr/Y ratios) for island-arc ADR volcanics and opposite to that for formation of adakite by partial melting of metamorphosed basalt (see the AFC trend in Fig. 2).

Figure 3, a 'spider' diagram of an adakite, a representative normal-type MORB and two partial-melting curves from a

cation; 10, Los Chocoyos, Guatemala (4 Myr)³⁶; 11, Mount St Helens (3 Myr)³⁷; 12, Baja of California, Mexico (1 Myr)³⁸; 13, Woodlark Basin (0.5 Myr)³⁹.

MORB and a MORB and sediment (assuming a hornblende-eclogite restite)¹³, shows that trace elements can be modelled by partial melting of oceanic crust. There are strong positive Sr and Eu anomalies and depletions in Y and Yb in both the adakite and modelled melts.

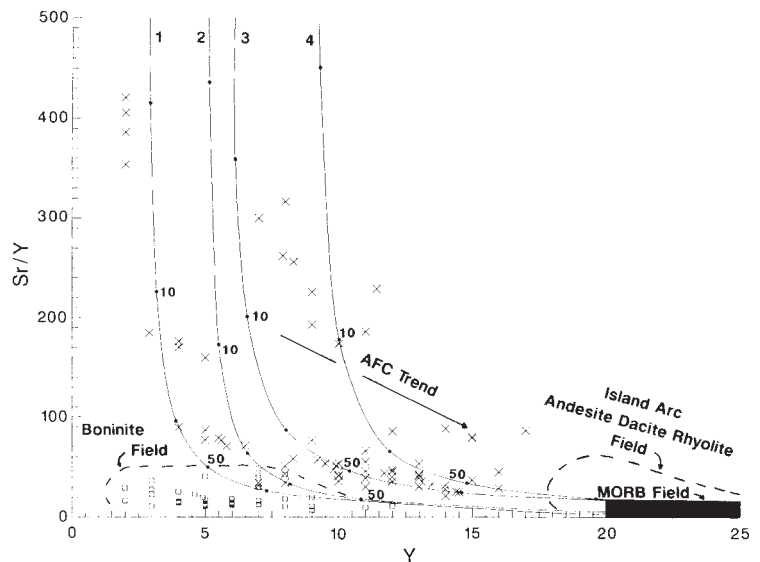
Although partial melting of subducted oceanic lithosphere seems to be the most internally consistent model for adakite generation, other types of petrogenesis are possible. The lower crust is a likely candidate because the geochemical signatures of adakites suggest a basaltic source. We contend, however, that partial melting of the lower crust by underplating will yield magmas with negative Eu anomalies and $\text{Al}_2\text{O}_3 < 15\%$ because basalt will not reach the amphibolite-eclogite transition under most arcs¹⁵. Similar geochemical signatures would be imparted to melts from plagioclase-rich granulite. Adakites have high $\text{Al}_2\text{O}_3 (> 15\%)$ and absent or positive Eu anomalies (Fig. 3).

Partial-melting experiments on high-alumina basalts (possible lower-crustal rocks) indicate that adakites are not products of underplating¹⁶. Low-alumina (12.3% Al_2O_3) dacitic melts were generated by 5–10% partial melting (16 kbar and $1,050^\circ\text{C}$ under amphibole dehydration conditions). In contrast, partial melting of MORBs under subduction pressure and temperature conditions (20 kbar and $1,100^\circ\text{C}$) produced high-alumina (17.2% Al_2O_3) dacitic material with higher degrees of partial melting (10–20%)¹⁶.

Some Mg-rich andesites are found in 5 of the 13 adakite localities (some as high as 6% MgO), but they are somewhat rare. Perhaps partial melting of the mantle wedge is the source of some adakites. Although adakites from a few locations have elevated Mg concentrations, they are not comparable to concentrations found in boninites (most boninites have $> 10\%$ MgO). Also, boninites have distinctive Sr/Y and Y values (Fig. 2) and are absent of plagioclase (adakites contain plagioclase). Differentiation of a boninite parent magma generates low-Al-type dacitic magmas (for example, the low-K rhyolites of Saipan¹⁷ or the quartz dacites from the Bonin Islands¹⁸), which are distinct from adakites (high-Al type).

Kay¹² has proposed a model that may explain the generation of the high-Mg subset of adakites: dacitic magmas generated by the melting of a subducted slab rise into the hot overlying mantle wedge where they re-equilibrate with peridotite under

FIG. 2 Sr/Y plotted against Y (adakites, $\times$; boninites, $\square$). The curves represent various models of the partial melting of depleted and altered MORB with an amphibolite or eclogite residue: 1, eclogite (gt/cpx=50/50); 2, garnet amphibolite (gt/am=10/90); 3, amphibole eclogite (am/gt/cpx=10/40/50); 4, garnet amphibolite (gt/am=10/90). Starting compositions for curves 1 and 2: Sr=141 p.p.m. and Y=21 p.p.m.; curves 3 and 4: Sr=264 p.p.m. and Y=38 p.p.m. (see ref. 3 for further details). gt, garnet; cpx, clinopyroxene; am, amphibolite.



the hydrous conditions generating andesitic compositions. This process would explain the high-Mg content, relative to other adakites, and Y and HREE depletions found among this subset of rocks. Although we believe that the high-Mg adakites can be derived by the partial melting of the subducted slab without reaction with the mantle, we do not rule out Kay's alternative proposal. There is a continuum from low- to high-MgO adakites indicating that the high-Mg varieties are not unique. At any rate, these high-Mg adakites are a relatively small proportion of the adakites (<10% have MgO values ≥ 5 wt%). Further, partial melting of the mantle cannot be applied to the majority of adakites because of their relatively low MgO values (>75% have MgO values <4 wt%).

Modelling of fractional crystallization¹³ cannot explain the somewhat unique geochemistries of adakites. For example, plagioclase fractionation drives melts toward low Sr and high Y (Fig. 2) based on mineral-melt partition coefficients. Only amphibole and clinopyroxene can push magmas toward high Sr/Y and low Y, but these values cannot be reached if plagioclase fractionation along with the clinopyroxene and amphibole. Extensive amphibole fractionation (60%) is necessary to produce adakite REE compositions but 60% fractionation is inconsistent with the amount of amphibole normally found in adakites. Removal of this much subaluminous hornblende

would create a high-peraluminous felsic magma, which is inconsistent with the slightly peraluminous to metaluminous compositions of the adakites. Clinopyroxene is absent in many adakites and, therefore, it is difficult to see how fractionation of this mineral could generate adakites. Similar results are obtained when AFC or magma mixing processes are modelled¹³.

We conclude that the most likely derivation of adakites is from the partial melting of the subducted slab. This is supported by the existence of adakites associated with amphibolite and eclogite in exposed subduction zones. A melange complex, consisting of blocks of greenschist, blueschist, metagabbro, rhyolite (adakite), amphibolite and eclogite, has been described from Baja California Sur, Mexico¹⁹ beneath an Upper Triassic ophiolite. The Catalina schist of southern California is a subduction-zone metamorphic terrane. The interaction of fluids at high-pressure and temperature conditions initiated partial melting of metamorphosed basalt to generate trondhjemitic melts of adakite compositions in migmatitic blocks from the Catalina schist^{20,21}.

We suggest that only oceanic crust younger than 25 Myr is hot enough to initiate melting of the slab. Parsons and Sclater²² have plotted heat flow units (HFU) against the age of the oceanic crust. Crust younger than 25 Myr has HFU in the range ~2.8–8, but crust older than this has fairly constant HFU of 1–2.5.

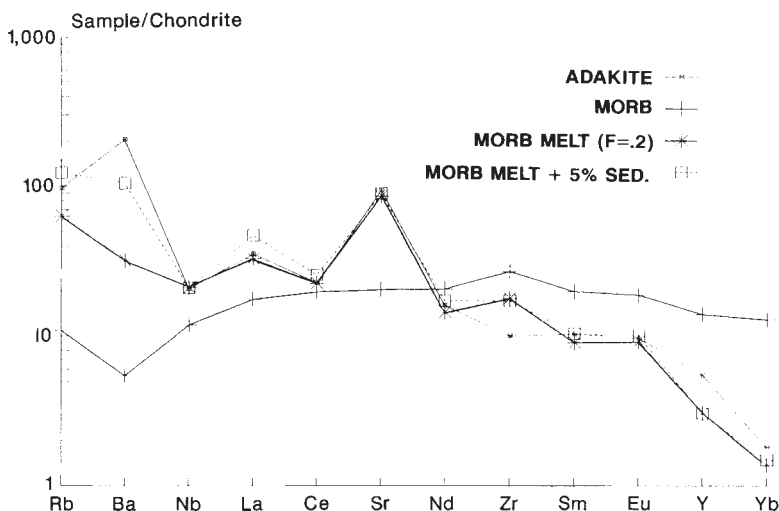


FIG. 3 Chondrite-normalized⁴⁰ compositions of a typical adakite and normal-type MORB and two models derived from 20% partial melting of the MORB and a mixture of 95% MORB and 2.5% marine sediments and/or 2.5% crustal sediments represented by an Archaean crustal component (see ref. 13 for details).

Perhaps subducted oceanic crust with HFU of at least 2.8 is needed to produce adakites during subduction. Based on phase diagrams and numerical heat-transfer models, Peacock²³ predicted that partial melting of the oceanic crust should occur "only within several tens of millions of years of the initiation of subduction in young oceanic lithosphere".

Archaean high-Al trondhjemites, tonalites and dacites have similar compositions to the adakites described above. The higher geothermal gradients in the Archaean probably led to the rapid generation and subduction of young hot crust. Partial melting of this subducted lithosphere may have been a dominant process in the production of dacitic and trondhjemitic compositions throughout the Archaean shield terranes^{3,24,25}.

Ellam and Hawkesworth²⁶ have suggested that "Either the continental crust is considerably more mafic than andesite, or the mechanism of crust formation has changed with time, such that a large proportion of the continental crust was formed by processes unlike those active in recent subduction zones." Adakites may be a modern analogue of Archaean magmatism. Their rarity in Cenozoic arcs can be related to the scarcity of subduction of young oceanic crust, explaining why the generation of 'typical' continental crust material is rare in arc regions today. □

Received 4 April; accepted 23 August 1990.

- Gill, J. B. in *Orogenic Andesites and Plate Tectonics* (Springer, Berlin, 1981).
- Defant, M. J. & Nielsen, R. L. *Geochim. cosmochim. Acta* **54**, 87–102 (1990).
- Drummond, M. S. & Defant, M. J. *J. geophys. Res.* (in the press).
- Cande, S. C. et al. *Magnetic Lineations of the World's Ocean Basins* (map) (Am. Ass. Petrol. Geol., 1989).

- Scotese, C. R., Gahagan, L. M. & Larson, R. L. *Tectonophysics* **155**, 27–48 (1989).
- Davidson, J. P. *Geochim. cosmochim. Acta* **51**, 2185–2198 (1987).
- Defant, M. J., de Boer, J. Z. & Oles, D. *Tectonophysics* **145**, 305–317 (1988).
- Kay, R. W. *J. Geol.* **88**, 497–522 (1980).
- Smith, I. E. M., Taylor, S. R. & Johnson, R. W. *Contr. Miner. Petrol.* **69**, 227–233 (1979).
- Gill, J. B. *Earth planet. Sci. Lett.* **68**, 443–458 (1984).
- Tatsumi, Y. & Koyaguchi, T. *Contr. Miner. Petrol.* **102**, 34–40 (1989).
- Kay, R. W. *J. Volcan. geotherm. Res.* **4**, 117–132 (1978).
- Defant, M. J. et al. *J. Petrol.* (submitted).
- DePaolo, D. J. *Earth planet. Sci. Lett.* **53**, 189–202 (1981).
- Tepper, J. H. *Eos* **70**, 1420 (1989).
- Rapp, R. P., Watson, E. B. & Miller, C. F. (abstr.) *Goldschmidt Conf. Prog.* 69 (Geochem. Soc., Baltimore, 1988).
- Meijer, A. *Contr. Miner. Petrol.* **83**, 45–51 (1983).
- Kuroda, N., Shiraki, K. & Urano, H. *Contr. Miner. Petrol.* **100**, 129–138 (1988).
- Moore, T. E. *Geol. Soc. Am. Mem.* **164**, 43–58 (1986).
- Sorenson, S. S. *Metamorph. Geol.* **6**, 405–435 (1988).
- Sorenson, S. S. & Grossman, J. N. *Geochim. cosmochim. Acta* **53**, 3155–3177 (1989).
- Parson, B. A. & Sclater, J. G. *J. geophys. Res.* **82**, 803–827 (1977).
- Peacock, S. M. *Science* **248**, 329–337 (1990).
- Jahn, B. M., Vidal, Ph. & Kröner, A. *Contr. Miner. Petrol.* **86**, 398–408 (1984).
- Martin, H. *Geology* **14**, 753–756 (1986).
- Ellam, R. M. & Hawkesworth, C. J. *Geology* **16**, 314–317 (1988).
- Masuda, A., Nakamura, N. & Tanaka, T. *Geochim. cosmochim. Acta* **37**, 239–244 (1973).
- Defant, M. J. et al. *Contr. Miner. Petrol.* (in the press).
- Puig, A., Herve, M., Suarez, M. & Saunders, A. D. *J. Volcan. geotherm. Res.* **20**, 149–163 (1984).
- Stern, C. R., Skewes, M. A. & Duran, M. *Actas Primer Congreso Geológico Chileno* **2**, 195–212 (1976).
- Stern, C. R., Futa, K. & Muehlenbachs, K. in *Andean Magmatism—Chemical and Isotopic Constraints* (eds Harmon, R. S. & Barreiro, B. A.) 31–46 (Shiva, Cheshire, 1984).
- Barker, F., Arth, J. G. & Stern, T. W. *Am. Miner.* **71**, 632–643 (1986).
- Kepezhinskas, P. K. *Geologiya i razvedka* **1**, 59–65 (1988).
- Cameron, K. L. & Cameron, M. *Contr. Miner. Petrol.* **91**, 1–11 (1985).
- Bartholomew, D. S. thesis, Univ. Leicester (1984).
- Rose, W. I., Grant, N. K. & Easter, J. in *Ash-Flow Tufts, Geol. Soc. Am., Spec. Pap.* 180, 87–99 (1979).
- Smith, D. R. & Leeman, W. P. *J. geophys. Res.* **92**, 10313–10334 (1987).
- Rogers, G., Saunders, A. D., Terrell, D. J., Verma, S. P. & Mariner, G. F. *Nature* **315**, 389–392 (1985).
- Johnson, R. W. et al. *Circum-Pacific Council for Energy and Mineral Resources, Earth Sci. Ser. Vol. 7* (eds Taylor, B. & Exon, N. F.) 155–226 (1987).
- Wakita, H., Rey, P. & Schmitt, R. A. *Proc. lunar Sci. Conf.* **2**, 1319–1329 (1971).

ACKNOWLEDGEMENTS. We thank U. Knittel for a review.

Transport of reef corals into the Great Barrier Reef

Paul L. Jokiel

University of Hawaii at Manoa, Hawaii Institute of Marine Biology,
PO Box 1346, Kaneohe, Hawaii 96744, USA

MAJOR submarine pumice eruptions have occurred repeatedly during the past 25 years in the Tonga–Kermadec region^{1–3}. Pumice from this area drifts through Fiji and New Caledonia, reaching Australia roughly one year after an eruption^{1,2}. Coral larvae settle and grow on the floating pumice fragments (Fig. 1). The geographic pattern of abundance and size distribution of rafted corals found on drift pumice in relation to the pumice source area suggests massive transport of corals into the Great Barrier Reef from regions of lower diversity lying far to the east and southeast. Dispersal range is not limited by longevity of larvae and centres of diversity may be sites of species accumulation rather than sites of species origin.

Mean colony diameter, maximum diameter, size range, standard deviation of the mean, and abundance of coral settlements found on drift pumice increase with increasing distance along the drift trajectory (Table 1 and Fig. 2). Differences in the distributions for each pair of locations are significant ($P < 0.001$ by the Kolmogorov–Smirnov two-sample test for distributions⁴), suggesting continuous growth of colonies after settlement with additional recruitment throughout the drift period. The largest rafted corals found in Australia have a diameter of ~2 cm. Corals of this size are about one year old^{5,6}, which is consistent with directly observed drift time for the pumice.

Corals of the family Pocilloporidae accounted for 70% of the total, with the genera *Pocillopora*, *Seriatopora* and *Stylophora* being represented. Acroporidae accounted for 25% of the total. The genus *Acropora* was abundant and several colonies of *Montipora* were clearly identifiable. The remaining 5% consisted

TABLE 1 Size and numbers of rafted corals collected per unit effort at various locations

Location	n	Mean $\pm$ s.d. (mm)	Range (mm)	Rafted corals collected	
				per h	per km
New Zealand	0	0	0	0	0
Fiji	15	1.5 $\pm$ 0.4	0.9–2.2	1	2
New Caledonia	12	2.9 $\pm$ 1.4	1.3–6.2	3	4
Australia	156	5.2 $\pm$ 4.8	0.5–22.0	5	10

of colonies of Poritidae (*Porites* and *Goniopora*) and Faviidae (*Favia* and *Cyphastrea*).

Pumice fragments that carried corals into the Great Barrier Reef have the distinctive elemental composition of pumice known to originate in the Tonga region (Table 2), and is unlike pumice that originates in other geographic regions^{1,7}. The major surface currents between Tonga and the Great Barrier Reef flow to the southwest (Trade Wind Drift), developing into a strong southerly current (East Australia Current) near the coastline of Australia, which eventually joins the Tasman Current flowing eastward to New Zealand^{8,9}. The observed drift trajectory of pumice actually approximates the pattern of surface wind in the region¹⁰, rather than the published generalized pattern of surface currents. Pumice drift direction is controlled by current direction rather than wind direction¹. The apparent contradiction is readily explained. Wind stress in the southeast Trade Wind region drags the shallowest layer of water, which contains the pumice, to the northwest. The deeper layers are increasingly deflected to the southwest with increasing depth due to the Coriolis effect and shear phenomena as described by the 'Ekman spiral'¹¹. Reported 'surface current' shown in generalized maps of net surface water transport actually represents integrated transport above the thermocline. The result is massive transport of corals into the Great Barrier Reef, suggesting that centres of high diversity may actually represent centres of species accumulation rather than centres of species origin^{12–14}.

The pelagic larvae of most inshore marine organisms typically have a lifespan of about one month¹⁵, as is the case for most reef corals^{16,17}. Studies of genetic population structure and recruitment¹⁸⁻²⁰ over distances of a few kilometres (km) suggest that dispersal by coral larvae may be extremely limited. Typical current velocity in the Fiji to the Great Barrier Reef region is about 20 cm s^{-1} , which would only carry larvae a distance of seven hundred kilometres over their lifespan of one month. Deep-water barriers greatly exceeding this distance are found between many island groups. Rafted corals and other organisms, however, can live for years and travel many thousands of kilometres.

Recruitment of coral larvae onto floating objects seems to be enhanced by the physical processes that sweep flotsam into 'windrows' or 'slicks'. Windrows are visually apparent when large amounts of drift pumice are present¹. Massive windrows of coral eggs, sperm and larvae often form during periods of mass spawning in waters of the Great Barrier Reef²¹. Objects drifting through reef areas during time of spawning will be driven into the windrows and brought into direct contact with the concentrated larvae (Fig. 1).

Pumice can float for years, but fouling by a coral will eventually cause it to sink (Fig. 1). If the pumice sinks in shallow

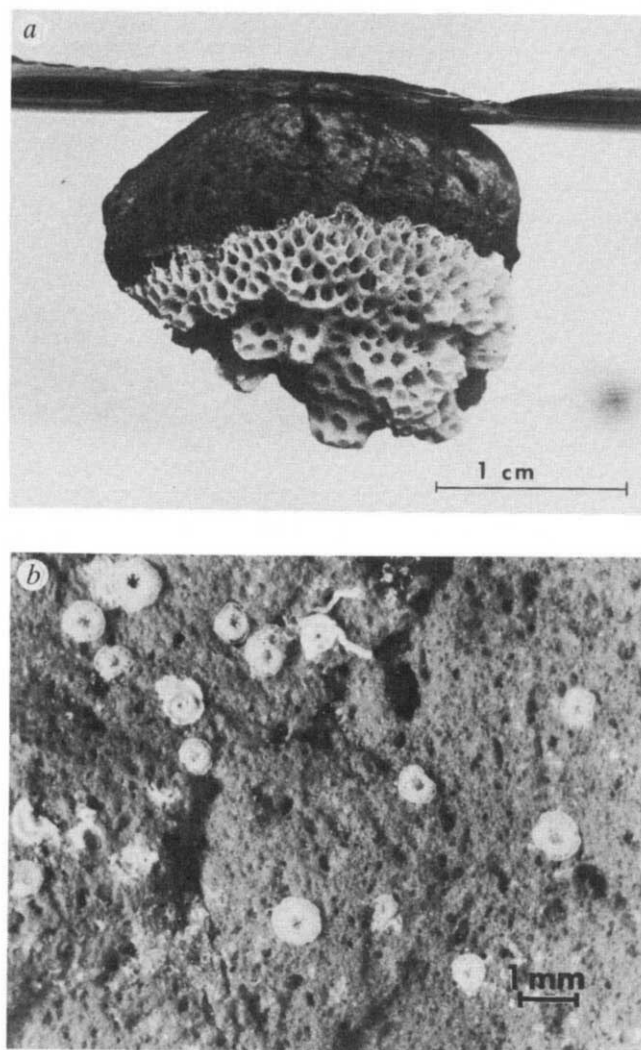


FIG. 1 *a*, Rafted skeleton of *Pocillopora* sp. from Kurramine Beach, Australia. Colony grew to a size that nearly sank the pumice before it was beached. *b*, Spat of *Acropora* sp. and serpulid worms on pumice collected on Strand Beach, Townsville, Australia in August 1988. The corals were probably derived from the annual mass spawning that occurred during June 1988.

TABLE 2 Elemental composition of pumice

	Pumice from eruptions in the Tonga region ⁷ range (%)	Pumice beached along the Great Barrier Reef	
		Mean $\pm$ s.d. (%)	Range (%)
TiO ₂	0.5–0.9	0.67 ± 0.15	0.66–0.70
Fe ₂ O ₃	5.3–10	8.48 ± 0.08	8.41–8.63
K ₂ O	0.5–1.0	0.87 ± 0.02	0.85–0.90
Na ₂ O/CaO	0.4–1.0	0.46 ± 0.00	0.45–0.46

Comparison between published⁷ values for elemental composition of pumice erupted in the Tonga area and coral-encrusted pumice beached along the Great Barrier Reef that was analysed on a Siemens SRS 303 X-ray fluorescence spectrometer.

water, the attached organisms could gain a foothold. Sunken drift-pumice could be an important source of volcanic material in pelagic sediments¹. Rafted corals may dislodge from a floating object upon impact with a reef and thereby become established. Also, some colonies grow to reproductive size while still attached to floating pumice^{12,22}.

Typical bulk pumice samples from the South Pacific gathered during this study contain $\sim 10^6$ fragments per m^3 . Large eruptions, such as the Home Reef eruption² of 1984 probably produce $\sim 10^8 \text{ m}^3$ of pumice, as estimated from reports of drift pumice and the temporary formation of an island measuring $0.5 \times 1.5 \text{ km}$. The total number of individual pumice floats produced in this single eruption was probably $\sim 10^{14}$. About 10 m^3 of pumice per km of shoreline is presently found along the Great Barrier Reef. Total length of shoreline containing beached pumice drift is $\sim 10^3 \text{ km}$. This survey yielded ~ 10 rafted colonies per km of shoreline on the Great Barrier Reef. Therefore, a minimum estimate for numbers of beached coral skeletons on pumice is of the order of about 10^4 colonies. This number probably represents a 10–20-year accumulation, yielding an approximate arrival rate of 10^3 colonies per year. This estimate is low because it does not include skeletons lost because of abrasion in beach drift, loss due to sinking, loss of colonies that passed through the region without being beached, or the number of additional colonies rafted into the area on organic materials such as logs, nuts and other floating debris.

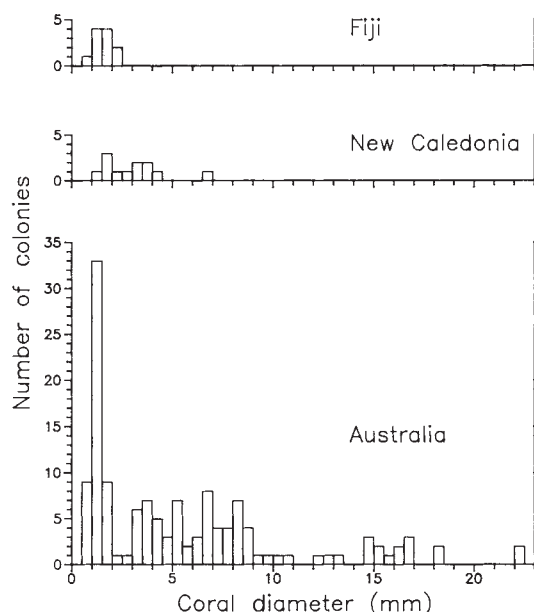


FIG. 2 Frequency (number of colonies) in each size category (colony diameter) at each location.

Pumice is abundant throughout the atolls of the Pacific^{23,24}, and is continually added through submarine and shoreline eruptions or land-derived pumice carried into the sea by rivers. Immense pumice eruptions have occurred throughout geological time. Several km³ of drift pumice were produced by the Krakatau eruption²⁵ of 1883. Pumice-rich pyroclastic rocks are abundant in marine formations throughout the geological record back to the Precambrian²⁶, so rafting of organisms on drift pumice has been an important mechanism of long-range dispersal throughout the history of life in the oceans. □

Received 15 May; accepted 25 July 1990.

- Bryan, W. B. *Geol. Soc. Am. Bull.* **82**, 2799–2812 (1971).
- Smithsonian Institution, Scientific Event Alert Network, *SEAN Bull.* **4**(5), 2–3 (1979); **4**(8), 2 (1979); **4**(10), 10 (1979); **4**(12), 6–8 (1979); **8**(4), 13 (1983); **9**(4), 7–8 (1984); **9**(7), 8–9 (1984); **9**(9), 5 (1984) Publication NTISUB/157, National Technological Information Service, US Department of Commerce, Springfield, Virginia, USA.
- Simkin, T. *et al. Volcanoes of the World* (Hutchinson Ross, Stroudsburg, Pennsylvania, 1981).
- Zar, J. H. *Biostatistical Analysis*, (Prentice-Hall, Englewood Cliffs, New Jersey, 1984).
- Grigg, R. W. & Maragos, J. E. *Ecology* **55**, 387–395 (1974).
- Polachek, T. thesis, Univ. of Hawaii, Honolulu (1978).
- Frick, C. & Kent, L. E. *Trans. geol. Soc. S. Afr.* **87**, 19–33 (1984).
- Pickard, G. L., Donguy, J. R., Henin, C. & F. Rougerie. *A Review of the Physical Oceanography of the Great Barrier Reef and Western Coral Sea*, (Australian Government Publishing Service, Canberra, 1977).
- Knox, G. A. *Scientific Exploration of the South Pacific* (ed. Wooster, W. S.) 155–182 (National Academy Science, Washington D.C., 1970).
- Ramage, C. S. *Scientific Exploration of the South Pacific* (ed. Wooster, W. S.) 16–29 (National Academy Science, Washington D.C., 1970).
- Sverdrup, H. U., Johnston, M. W. & Fleming, R. H. *The Oceans* (Prentice-Hall, New York, 1942).
- Jokiel, P. L. *Mar. Biol.* **101**, 483–493 (1989).
- Ladd, H. S. *Am. J. Sci.* **258**, 137–150 (1960).
- Kay, E. A. in *Biogeography of the Tropical Pacific* (eds Radovsky, F. J., Raven, P. H. & Sohmer, S. H. 15–31 (Bishop Mus. Special Pub. No. 72, Honolulu, 1984).
- Thorson, G. *Oceanography* (ed. Sears, M.) 455–474 (American Association for the Advancement of Science, Washington, D.C., 1961).
- Harrison, P. L., Babcock, R. C., Oliver, J. K., Wallace, C. C. & Willis, B. L. *Science* **223**, 1186–1189 (1984).
- Richmond, R. *Proc. Sixth Int. Symp. Coral Reefs* **2**, 827–831 (1988).
- Stoddart, J. A. *Coral Reefs* **3**, 149–156 (1984).
- Stoddart, J. A. *Aust. J. mar. Freshwat. Res.* **39**, 109–115 (1988).
- Sammacco, P. W. & Andrews, J. C. *Science* **239**, 1442–1424 (1988).
- Willis, B. K. & Oliver, J. K. *Proc. Sixth Int. Coral Reef Symp.* **2**, 853–859 (1988).
- Jokiel, P. L. *Coral Reefs* **3**, 113–116 (1984).
- Sachet, M.-H. *Atoll Res. Bull.* **37**, 1–27 (1955).
- Wiens, H. J. *Atoll Environment and Ecology* (Yale University Press, London, 1962).
- Simkin, T. & Fiske, R. S. *Krakatau 1883* (Smithsonian Institution Press, Washington D.C., 1983).
- Fiske, R. S. *Geol. Soc. Am. Bull.* **80**, 1–8 (1969).

Embryonic expression of a haematopoietic growth factor encoded by the *Sl* locus and the ligand for c-kit

Yasuhisa Matsui*, Kristina M. Zsebo†
& Brigid L. M. Hogan*‡

* Department of Cell Biology, Vanderbilt University Medical School, Nashville, Tennessee 37232-2172, USA

† Amgen Incorporated, Amgen Center, Thousand Oaks, California 91320, USA

MICE carrying mutations at the *W* (Dominant white spotting) and *Sl* (Steel) loci develop abnormalities in three independent systems: neural crest-derived melanocytes, primordial germ cells and haematopoietic stem cells. Consequently, homozygotes of viable mutant alleles have white coats and are sterile and severely anaemic. Tissue recombination studies predict that the *W* gene is expressed cell autonomously, whereas the product of the *Sl* locus affects the microenvironment in which the stem cells migrate, proliferate and differentiate^{1–6}. The *W* locus encodes the proto-oncogene c-kit, a member of the tyrosine kinase receptor family^{7,8}.

The haematopoietic growth factor SCF (stem cell factor) has been identified as the product of the *Sl* locus and a ligand for c-kit^{9–11}. Here, we report that SCF is expressed during embryogenesis in cells associated with both the migratory pathways and homing sites of melanoblasts, germ cells and haematopoietic stem cells. Both SCF and c-kit are also expressed in a variety of other tissues, including the brain and spinal cord, suggesting that the receptor–ligand system has additional roles in embryogenesis.

The first indication that SCF has a diverse pattern of expression came from northern blot analysis of RNA isolated from embryonic and adult mouse tissues. A transcript of ~6.2 kilobases (kb) is present in embryos from 10.5–17.5 days post coitum (p.c.) and in organs associated with haematopoiesis (yolk sac, embryonic liver) and germ cells (ovary, testis). In addition, SCF transcripts are detected in total RNA from brain, lung, kidney and a variety of other organs. To characterize this expression further, we analysed the spatial distribution of SCF RNA by *in situ* hybridization.

As would be expected for a haematopoietic growth factor, SCF transcripts are detected in the yolk sac (endoderm and mesoderm), embryonic liver, and developing bone (Fig. 2a, b, j, k, o, p, q, r, and data not shown). But even at early stages of development, the hybridization signal in the yolk sac is much lower than in the fetal liver, a finding consistent with the northern analysis (Fig. 1). SCF expression in bone is not confined to the central region where haematopoiesis takes place, but is also seen in the densely packed cells of the outer periosteal layer (see, for example, Fig. 2q, r), suggesting an additional role for the gene product in osteogenesis.

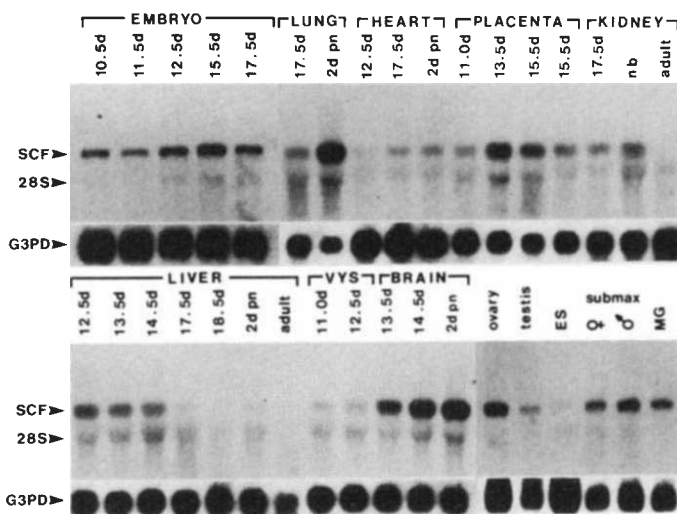


FIG. 1 Northern analysis of SCF gene transcripts in mouse embryos and embryonic and adult tissues. RNA blots were prepared using either 5 µg poly(A)⁺ RNA from total embryos of 10.5–17.5 days p.c. (noon of day of plug is 0.5 days p.c.) and adult tissues or 20 µg total RNA from the indicated embryonic tissues, and hybridized sequentially with the murine SCF probe and a murine probe for glyceraldehyde-3-phosphate dehydrogenase (G3PD). An SCF transcript of ~6.2 kb was detected in many tissues. Some cross-hybridization with 28S ribosomal RNA was also observed in total RNA samples. Abbreviations: VYS, visceral yolk sac; ES, undifferentiated embryonic stem cells; submax, submaxillary gland; MG, mammary gland; d, days p.c.; pn, days post natal; nb, newborn.

METHODS. RNA was prepared by the LiCl-urea method and electrophoresed in 1% agarose-formaldehyde gels, transferred to Zetabind membranes and baked. The antisense SCF RNA probe was prepared using T7 RNA polymerase from a 396-bp complementary DNA encoding amino acids 29–160 (ref. 9) of the mouse protein inserted into the pGEM7Z(+) vector. The 280-bp G3PD probe has been described¹⁸. Filters were hybridized under the following conditions: 5 × SSCP buffer, 5 × Denhardt's solution, 50 µg ml⁻¹ sonicated salmon sperm DNA, 0.1% SDS, 50% formamide at 65 °C for 15 h, then washed in 0.2 × SSCP, 0.1% SDS at 65 °C for 1 h.

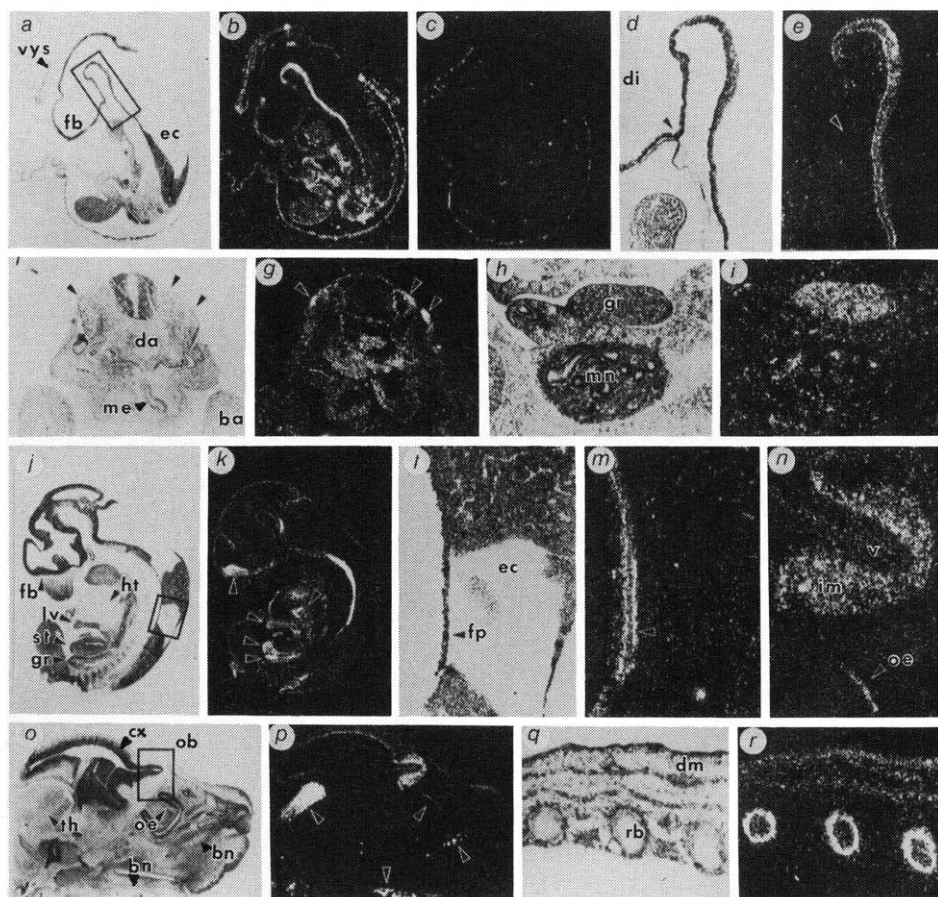
‡ To whom correspondence should be addressed.

Mice homozygous for many viable *Sl* mutations are sterile because of a failure in the proliferation of primordial germ cells and in their migration from the base of the allantois to the genital ridges^{2,6}. From about 9.0 days p.c., SCF transcripts are clearly present in mesodermal cells along the migratory pathway of the primordial germ cells, for example around the hind gut and the ventral wall of the dorsal aorta, and in the mesentery. Expression is also seen in the genital ridges themselves (Fig. 2f–k, and data not shown). Mesodermal expression is not limited to these regions, and includes mesonephric and metanephric kidney tubules (Fig. 2h, i), mesoderm of the stomach (Fig. 2j, k) and lung, the atrioventricular cushions of the heart (Fig. 2j, k) and the regions within the neural crest-derived mesenchyme of the facial processes and ventral branchial arches (data not shown). Analysis of the genital ridges by high-power microscopy showed that the hybridization grains are evenly distributed over stromal cells and are not localized in the germ cells, which can be distinguished by their round shape, larger size and prominent nuclei. In contrast, transcripts of the *c-kit* gene, which encodes a receptor for SCF, have been detected only in the primordial germ cells of the genital ridge, and not in the stroma (ref. 12, and R. Bachvarova, personal communication).

Mutations at the *Sl* locus characteristically affect another

migratory stem cell population—the neural crest-derived melanocytes. In the trunk and tail of the mouse, precursors of the melanoblasts migrate between the ectoderm and the dorsal surface of the somites, whereas the precursors of the sympathetic nervous system take more ventral routes¹³. Analysis of transverse sections of 9.5- and 10.5-days p.c. embryos, and the posterior parts of later embryos, reveals high levels of SCF expression in the dorsal region of the somites, over which the presumptive melanoblasts migrate (Fig. 2f, g, arrows), and much lower levels in the ventrolateral and ventromedial routes of neural crest migration. In addition, SCF transcripts are present in the sites to which the melanoblasts migrate, for example the ectoderm of whisker follicles at 12.5-days p.c., around the otic vesicle and developing ear, and in the dermis (but not the epidermis) of the skin (Fig. 2q, r). High-power microscopy shows that the hybridization grains over the dorsal somite are localized in the epithelial presumptive dermamyotome cells and not in neural crest cells *en route* to the extremities. SCF RNA is also absent from neural crest-derived dorsal root and cranial ganglia (data not shown). In contrast, *c-kit* RNA has been detected in presumptive melanoblasts in the dermis of 10.5-day p.c. embryos, and in the dorsal root ganglia, but is absent from the somites (ref. 12 and R. Bachvarova, personal communication).

FIG. 2 Spatial localization of SCF gene transcripts revealed by *in situ* hybridization analysis of sections of mouse embryos. a–e, Photomicrographs of sagittal sections through a 9.5-day p.c. embryo hybridized with antisense (a, b, d, e) or sense (c) strand SCF riboprobe. d and e, Higher magnification views of the boxed region in (a) showing SCF transcripts in the floor plate of the mid and hind brain. Reconstruction of serial sections showed that this expression extends anteriorly to the level of the developing infundibulum (arrow), that is, to the anterior limit of the floor plate. f and g, Transverse sections through a 10.5-day p.c. embryo showing SCF expression in the dorsal region of the somites (arrows), in the mesoderm of the ventral wall of the dorsal aorta, and in the mesentery of the hind gut. Because of the angle of sectioning, two somites are cut at different levels on the left side of the embryo but only on the right. The apparent signal from the red blood cells in the dorsal aorta is an artefact. h, i, Transverse section through the genital ridge and metanephros of a 12.5-day p.c. embryo. j–m, Sagittal section through an 11.5-day p.c. embryo showing SCF transcripts in the forebrain, mesoderm of the genital ridge, stomach, liver and cushions of the atrioventricular canals and truncus arteriosus of the heart. l and m, Higher magnification views of boxed region in (j) showing SCF transcripts in the floor plate of the anterior spinal cord. n, Higher magnification view of the olfactory bulb region boxed in (o). Hybridization grains are present over both the ventricular and intermediate cell layers, and in the olfactory epithelial cells of Jacobson's organ. o and p, Sagittal section through the head of a 14.5-day p.c. embryo showing SCF transcripts in the olfactory bulbs, thalamus, penultimate outer layer of the cortex and bone. q, r, Detail of a sagittal section through a 14.5-day p.c. embryo showing SCF transcripts in the dermis of the skin and in the periosteal layer around the developing ribs. Adjacent sections hybridized with sense strand probe gave only very low levels of hybridization grains around the ribs and background levels in the dermis. Scale bars, 0.5 mm in a, f, j, o and 0.1 mm in q. Abbreviations: ba, branchial



arch; bn, bone; cx, cortex; da, dorsal aorta; di, diencephalon; dm, dermis; ec, ependymal canal; fb, forebrain; fp, floorplate; gr, genital ridge; ht, heart; im, intermediate layer; me, mesentery; mn, metanephros; ob, olfactory bulb; oe, olfactory epithelium; rb, rib; st, stomach; th, thalamus; v, ventricular layer; vys, visceral yolk sac.

METHODS. *In situ* hybridization using ³⁵S-labelled single-strand riboprobes and high stringency conditions was carried out as described¹⁹, except that the hybridization and washing temperatures were 55 °C and 65 °C, respectively. The sense and antisense probes were synthesized from the same 396-bp SCF cDNA insert in pGEM7zf(+) as used for Fig. 1.

The SCF gene is clearly expressed in tissues associated with the migration, proliferation and differentiation of haematopoietic stem cells, presumptive melanoblasts and primordial germ cells, but transcripts are also present in a variety of other sites. In particular, a strong hybridization signal for SCF is found in the spinal cord and brain (Fig. 2*a, b, d, e, j-p*). Reconstruction of serial sections of 9.5- and 10.5-day p.c. embryos hybridized with the SCF probe showed that the RNA is specifically localized in the midline floorplate cells, extending anteriorly to the diencephalon and terminating at about the level of the developing pituitary. SCF transcripts are also detected in the forebrain from around 9.5 days p.c., in cells destined to give rise to the olfactory bulbs (Fig. 2*a, b, j, k*). By 14.5 days, SCF transcripts are in both the ventricular and intermediate layers of the olfactory bulbs (Fig. 2*n-p*), a pattern that persists in the newborn. SCF transcripts are also present in the cells of the olfactory neuroepithelium, presumably including sensory neurons that innervate the olfactory bulbs. Complete analysis has not yet been made of SCF expression in the late gestation and newborn mouse brain, but preliminary studies detect transcripts in specific regions of the cerebellum, temporal lobes and thalamus (Fig. 2*o, p* and data not shown). Other investigators have reported high levels of *c-kit* RNA in the developing central nervous system (CNS) of the mouse, including the dorsal spinal cord and brain (cerebellum, hippocampus and telen-cephalon)^{12,14}. Taken together, these results invite speculation that the *c-kit* protein receptor-SCF ligand system has important roles in cell migration, proliferation and differentiation in the CNS. For example, floorplate cells have been shown to specifically produce a diffusible chemotropic factor(s) that promotes and directs the outgrowth of commissural axons originating in the dorsal spinal cord¹⁵⁻¹⁷. The high level of SCF RNA found in the floor plate cells raises the possibility that the protein is synthesized and secreted by these cells during the time when commissural tracts are being established. Arguing against a role for *c-kit*-SCF in the developing CNS is the fact that no gross neurological defects have so far been reported in homozygous *Sl/Sl* embryos, which survive *in utero* until ~15-16 days p.c., even though in these mutants the *Sl* gene is apparently deleted¹¹. It is, of course, possible that the neuronal transcripts of SCF and *c-kit* are untranslated, or that there is sufficient redundancy in receptor-ligand molecules in the CNS to allow deletion of one signalling component without phenotypic effect. Alternatively, subtle changes in the number and organization of specific neuronal or glial populations may have gone undetected. These alternatives are currently under investigation. □

Selective attraction of monocytes and T lymphocytes of the memory phenotype by cytokine RANTES

Thomas J. Schall*, Kevin Bacon†, Karen J. Toy* & David V. Goeddel*

* Genentech Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, USA

† Institute of Dermatology, St Thomas's Hospital, Lambeth Palace Road, London SE1 7EH, UK

AN important process in the immune response is the migration of different populations of lymphocytes at the proper time to sites of antigenic challenge. Although several chemoattractants are known for broad classes of lymphocytes, such as T and B cells, the process by which lymphocytes of specific subsets, such as helper, cytotoxic or memory T cells, migrate to the appropriate sites remains obscure. Interleukin-8 is a chemoattractant for T cells and neutrophils and is a member of a superfamily of soluble molecules related by a conserved motif containing four cysteine residues. IL-8 and related molecules, including platelet factor 4, constitute the C-X-C class of the superfamily and a group of cytokines produced by haematopoietic cells^{3,4} constitute the RANTES/sis or C-C class^{4,5}. The roles of most of these molecules are not well known, although murine MIP-1 α of the C-C branch is a specific inhibitor of haematopoietic stem cell proliferation⁶ and some members of the C-X-C branch are neutrophil-targeted inflammatory agents⁷⁻¹⁰. Here we report that the RANTES protein³ of the C-C class causes the selective migration of human blood monocytes and of T lymphocytes expressing the cell surface antigens CD4 and UCHL1. CD4⁺/UCHL1⁺T cells are thought to be prestimulated or primed helper T cells involved in memory T cell function¹¹⁻¹⁴. The preferential attraction of T-cell subsets by specific cytokines could in part explain how lymphocytes are targeted, and may provide insight into the workings of T cell memory.

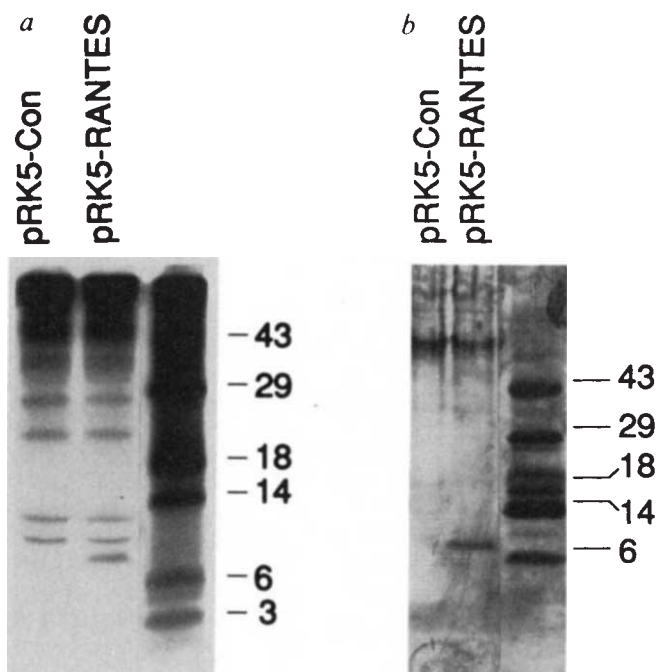
To investigate the properties of RANTES, we obtained conditioned media containing recombinant human RANTES using a mammalian cell expression system. The coding region of RANTES complementary DNA was cloned into the mammalian expression vector pRK5 under the control of the cytomegalovirus immediate early promoter/enhancer and simian virus 40 (SV40) termination and polyadenylation signals to give the vector pRK5-RANTES. The cDNA for RANTES was originally isolated from a subtracted (human T cell minus B cell) cDNA library and is predicted to encode a basic protein of relative molecular mass 8,000 (M_r 8K) (ref. 3). Transcription of RANTES mRNA seems to be markedly reduced in functional T cells after antigenic stimulation³. The human embryonic kidney cell line 293 (ref. 15) was transfected with pRK5-RANTES using the calcium phosphate precipitation method¹⁶. Transfected cells were metabolically labelled with [³⁵S]cysteine and [³⁵S]methionine, and the cell culture supernatants collected and analysed by Tricine-Tris SDS-PAGE. A band corresponding to 8K was observed in the supernatants from the 293 cells transfected with pRK5-RANTES, but not in the supernatants of control cells transfected with pRK5 (Fig. 1*a*). Also, an 8K band was detected by western blotting with antisera raised against RANTES N-terminal and C-terminal peptides (Fig. 1*b*). No band was detected in control cells transfected with pRK5. The 8K band therefore represents the recombinant RANTES cytokine.

To test the effect of recombinant RANTES on the migration of human blood monocytes and neutrophils, we performed a series of Boyden chamber assays on these cells. The results show that RANTES is a potent attractor of blood monocytes (Fig. 2). In each test, the undiluted RANTES-conditioned medium

Received 10 September; accepted 24 September 1990.

1. Sarvella, P. A. & Russell, L. B. *J. Hered.* **47**, 123-128 (1956).
2. Mintz, B. & Russell, E. S. *J. exp. Zool.* **134**, 207-230 (1957).
3. Mayer, T. C. & Green, M. C. *Dev. Biol.* **18**, 62-75 (1968).
4. Russell, E. S. *Adv. Genet.* **20**, 357-459 (1979).
5. Silvers, W. K. in *The Coat Colors of Mice: A Model for Gene Action and Interaction* 206-241 (Springer, New York, 1979).
6. Green, M. C. in *Genetics Variants and Strains of the Laboratory Mouse* (Eds Lyon, M. F. & Searle, A. G.) 333-335 (Oxford University Press, 1989).
7. Chabot, B., Stephenson, D. A., Chapman, V. M., Besmer, P. & Bernstein, A. *Nature* **335**, 88-89 (1988).
8. Geissler, E. N., Ryan, M. A. & Housman, D. E. *Cell* **55**, 185-192 (1988).
9. Zsebo, K. M. *et al. Cell* **63**, 195-201 (1990).
10. Martin, F. H. *et al. Cell* **63**, 203-211 (1990).
11. Zsebo, K. M. *et al. Cell* **63**, 213-224 (1990).
12. Orr-Urtreger, A. *et al. Development* **109**, 911-923 (1990).
13. Serbedzija, G. N., Fraser, S. E. & Bronner-Fraser, M. *Development* **108**, 605-612 (1990).
14. Nocka, K. *et al. Genes Dev.* **3**, 816-826 (1989).
15. Tessier-Lavigne, M., Placzek, M., Lumsden, A. G. S., Dodd, J. & Jessell, T. M. *Nature* **336**, 775-778 (1988).
16. Dodd, J. & Jessell, T. M. *Science* **242**, 692-699 (1988).
17. Jessell, T. M., Bovolenta, P., Placzek, M., Tessier-Lavigne, M. & Dodd, J. *Ciba Fdn Symp.* **144**, 255-280 (1989).
18. Lyons, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 4554-4558 (1989).
19. Pelton, R. W., Nomura, S., Moses, H. L. & Hogan, B. L. M. *Development* **106**, 759-767 (1989).

ACKNOWLEDGEMENTS. We thank F. Martin, R. Jacobsen and S. Suggs of Amgen, Inc. for the SCF cDNA probe, R. Bachvarova (Cornell Medical College) and P. Lonai (The Weizmann Institute) for communicating their results before publication, R. Pelton and M. Jones for help with photography, N. Wall for identifying brain parts, and C. Wright, K. Lyons and M. Jones for helpful comments on the manuscript.



collected between 4 and 6 h after the addition of serum-free medium to the transfected cells was considerably more effective in attracting monocytes than the positive control chemoattractant peptide, *N*-formylmethionine-leucyl-phenylalanine (fMLP), at concentrations of 10^{-8} or 10^{-9} M. By contrast, supernatants from control cells transfected with pRK5 showed little monocyte chemoattractant activity. RANTES-conditioned medium attracted only as many blood neutrophils as RPMI medium alone (Fig. 2, Con 1), although the control medium seemed slightly to inhibit the random migration of neutrophils in the assay (Fig. 2, Con 2). So although RANTES could have some attractive effect on neutrophils, its action seems largely preferential for blood monocytes. The ability of RANTES to attract monocytes in preference to neutrophils is in contrast to interleukin-8 (IL-8), which attracts neutrophils but not monocytes⁸.

The ability of medium conditioned with recombinant RANTES to attract lymphocytes isolated from peripheral blood was tested using a 48-well microchemotaxis chamber¹⁷. Leukocytes, isolated from whole blood, were depleted of monocytes by adherence to plastic for 18–24 h. The resulting cell distribution, as evaluated immunocytochemically^{17,18}, was <1% monocytes, 11% HLA-DR-bearing cells, 19% B cells and 79% T cells, of which 32% were CD8⁺, 67% CD4⁺, 38% UCHL1⁺(CD45RO) and 60% SN130⁺(CD45RA). RANTES-conditioned and control media were tested in parallel in duplicate wells of the microchemotaxis chamber. Lymphocyte migration was evaluated using an image analyser which measured the area of the lower surface of the filter occupied by cells¹⁷. The migration index (MI) is the area of lower surface of filter occupied by cells after stimulation with chemoattractant divided by the area of filter occupied by unstimulated randomly migrating cells. RANTES-conditioned medium had a significant and titratable migration index (MI = 2.3 at a 1/256 dilution), whereas control medium had no significant index (Fig. 3a). The MI for diluted medium conditioned with RANTES was roughly that of the positive control, 5% zymosan activated plasma (ZAP).

Migrating lymphocytes were further characterized by using a panel of monoclonal antibodies (legend to Fig. 3b) and by APAAP (alkaline phosphatase anti alkaline phosphatase) staining. There was marked migration of T cells, but not B cells, in response to RANTES (Fig. 3b). The distribution of T-cell

FIG. 1 Analysis of conditioned media from transfected 293 cells. *a*, Tricine SDS-PAGE of biosynthetically labelled supernatants from cells transfected with either plasmid pRK5-RANTES or pRK5 without a cDNA insert (pRK5-Con) fractionated on a 15% Tricine-Tris SDS polyacrylamide gel. A corresponding band to ~8K is seen specifically in the supernatants from cells transfected with plasmid pRK5-RANTES. Right-hand lane is migration of M_r markers (BioRad ¹⁴C low M_r standards; $M_r \times 10^{-3}$). *b*, Western blotting of transfected cell-conditioned media using RANTES antipeptide antisera. Pooled antisera generated against synthetic peptides representing the 10 N-terminal amino acids and the 19 C-terminal residues of the predicted RANTES mature protein specifically stain an 8K band in the supernatants of cells transfected with pRK5-RANTES. Right-hand lane is migration of M_r markers (BRL pre-stained low M_r standards; $M_r \times 10^{-3}$).

METHODS. RANTES cDNA containing the coding region (base pairs 1–411, ref. 3) was cloned into pRK5 (R. Klein and D.V.G., unpublished data) and used to transfect the human embryonic kidney cell line 293 (ref. 15) as previously described (ref. 16) using 10 μ g DNA per 100-mm plate. Control 293 cells were transfected with pRK5 containing no cDNA insert. A portion of the transfected cells was biosynthetically labelled with [³⁵S]methionine and [³⁵S]cysteine as previously described¹⁶. Supernatants were collected 6 h after the addition of serum-free medium and electrophoresed on a 15% Tricine-Tris SDS polyacrylamide gel²³. The gel was then fixed, soaked in autoradiography-enhancing solution (DuPont Enlightening) dried at 60 °C, and put directly on film overnight at -70 °C. Antisera for western blotting was generated by immunizing female New Zealand white rabbits with the synthetic peptides RAN01 (SPYSSDTPC single letter amino-acid code) or RAN02 (CANPEKKWVREYINSLEMS; single-letter amino-acid code) representing the N and C terminus of the mature RANTES protein, respectively; these peptides were conjugated to soybean trypsin inhibitor by means of their cysteine residues. Preimmune serum was collected from rabbits before immunization. Immunologic blotting performed as previously described²⁴. A duplicate blot was treated identically except it was incubated in pooled preimmune serum.

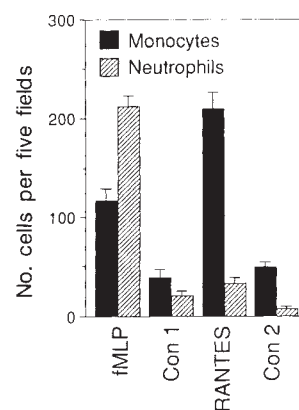


FIG. 2 Monocyte and neutrophil migration in response to RANTES. Results of experiments showing the migration of fresh blood monocytes (solid bars) and neutrophils (hatched bars). Test substances were fMLP peptide optimally diluted in RPMI medium, RPMI medium alone (Con 1), conditioned medium collected from 293 cells transfected with pRK5 (Con 2), or with pRK5-RANTES (RANTES). Histograms represent number of cells counted (mean $\pm$ s.e.m.) in five 40 $\times$ microscope objective fields.

METHODS. Human mononuclear and polymorphonuclear cells were isolated from heparinized venous blood of normal volunteers. Chemotaxis was assayed in a 48-well microchemotaxis chamber (Neuro Probe, Inc.) using the manufacturer's protocol. The lower compartments of the microchemotaxis chamber contained 100 μ l of serial 10-fold dilutions (10^{-6} M to 10^{-9} M) of fMLP (Sigma) in RPMI 1640 with L-Glutamine, 25 mM HEPES and 0.1 BSA, RPMI solution without fMLP as a negative control, or the transfected cell-conditioned media to be tested. Conditioned media consisted of cell culture supernatants collected 4 to 6 h after addition of serum-free medium to 293 cells transfected with either the control plasmid or with pRK5-RANTES. All samples were tested in triplicate wells. Results represent five experiments for monocytes and three for neutrophils. Values shown for fMLP response are concentrations (10^{-8} M or 10^{-9} M) causing the optimal response in each assay. Results were replicated with separate preparations of conditioned media on separate cell preparations.

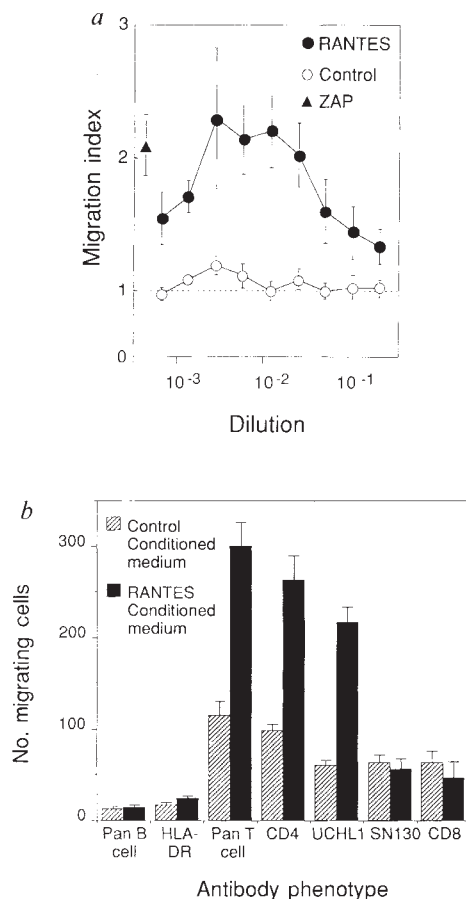


FIG. 3 Peripheral blood lymphocytes (PBL) migration in response to RANTES. **a**, Dose-response curves for lymphocyte migration to RANTES-conditioned medium (●) or control conditioned medium (○). Points represent serial two-fold dilutions of the conditioned media. Positive control is 5% zymosan-activated plasma (ZAP, ▲). Migration index defined as area of lower surface of filter occupied by stimulated cells divided by area of lower surface of filter occupied by unstimulated cells. Dotted line indicates value for randomly migrating cells. Results are expressed as mean migration indices with error bars representing s.e.m. for six experiments. **b**, Phenotype of lymphocytes migrating in response to RANTES (solid bars) or control media (hatched bars) as determined immunocytochemically. The phenotypes of PBL assayed as in **a** were determined using primary antibodies recognizing B cells, major histocompatibility complex class II surface antigens, all T cells, or the T-cell subset-specific markers CD4, CD8, UCHL1 or SN130. Histograms represent the mean number $\pm$ s.e.m. of positively staining cells which have migrated to the undersurface of the filters as counted per five high power fields. **METHODS**. Lymphocyte migration in response to RANTES was assessed using mixed PBL containing <1% monocytes in a 48-well microchemotaxis chamber, and quantified by automated image analysis as previously described^{2,17}. Each dose was tested in duplicate wells in the microchemotaxis chamber. After incubation for 1 h, phenotypes of the lymphocyte subsets migrating in response to RANTES-conditioned medium were compared with those migrating in the presence of control medium. The phenotypes of migrating lymphocytes were determined immunocytochemically using the protocol previously described for monocytes¹⁷ except that a 1:1 acetone:methanol mixture was used to fix the cells. The primary antibodies used were Dako T4 (1/200 dilution) Dako T8 (1/800 dilution) (Dako, High Wycombe, UK), Royal Free Pan T (undiluted) and Royal Free Pan B (undiluted) (Academic Department of Immunology, Royal Free Hospital, London, UK), UCHL1 (undiluted); P. C. Beverley, University College and Middlesex School of Medicine, London, UK) and SN130 (undiluted); L. Poulter, Royal Free Hospital Medical School, London, UK).

phenotypes was striking; there was no significant migration of CD8⁺ or SN130⁺ T cells in response to RANTES. Rather, migration is limited to CD4⁺ and UCHL1⁺ cells.

UCHL1⁺ CD4⁺ T helper cells are associated with memory T cell functions. Memory or immunologically experienced T cells express the UCHL1 antigen (CD45RO), which is the p180 isoform of the leukocyte common antigen CD45. By contrast, naive T cells express the p220/205 isoforms of CD45, designated CD45RA, which is recognized by the SN130 monoclonal antibody¹¹. Memory T cells are thought to be long-lived cells that continuously recirculate between blood and lymphoid tissues. On antigen restimulation of these cells, the memory response occurs probably as a result of their rapid clonal expansion. Also, memory T cells are more potent effectors than naive T cells because they are more able to produce certain multipotent cytokines on antigenic restimulation¹⁹⁻²¹. Memory T cells seem to partition preferentially into blood and tissues²², and are probably involved in chronic autoimmune disorders mediated

by T cells.

IL-8 is chemotactic for T lymphocytes^{1,2}. It attracts T cells at 100-fold lower concentrations than those necessary for neutrophils, and RANTES attracts T cells at 200-fold lower concentrations than the optimal monocyte concentration. But the activities of IL-8 and RANTES are distinct in this system, as IL-8 attracts all T-cell phenotypes tested equally (J. S. Ross, personal communication). Although belonging to the same superfamily as RANTES, IL-8 attracts neutrophils but not monocytes, whereas RANTES attracts monocytes but not neutrophils. Other members of the C-C branch do not seem specifically to attract memory T cells and have distinct activities (T.S. & K.B. unpublished results).

Our studies show that the recombinant RANTES cytokine can attract blood monocytes as well as T cells of the memory phenotype *in vitro*. These properties make it a candidate for roles in important immune regulatory functions, as well as autoimmune disorders and inflammatory processes. □

Received 16 July; accepted 9 August 1990.

1. Larsen, C. G., Anderson, A. N., Appella, E., Oppenheim, J. J. & Matsushima, K. *Science* **243**, 1461-1466 (1989).
2. Bacon, K. B., Westwick, J. & Camp, R. D. *Biochem. Biophys. Res. Commun.* **165**, 349-354 (1989).
3. Schall, T. J. *et al.* *J. Immunol.* **141**, 1018-1025 (1988).
4. Brown, K. D., Zurawski, S. M., Mosmann, T. R. & Zurawski, G. *J. Immunol.* **142**, 679-687 (1989).
5. Zipfel, P. F., Balke, J., Irving, S. G., Kelly, K. & Siebenlist, U. *J. Immunol.* **142**, 1582-1590 (1989).
6. Graham, G. J. *et al.* *Nature* **344**, 442-444 (1990).
7. Matsushima, K. *et al.* *J. exp. Med.* **167**, 1883-1892 (1988).
8. Yoshimura, T. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 9233-9237 (1987).
9. Walz, A., DeWald, B., von Tschanner, V. & Baggiolini, M. *J. exp. Med.* **170**, 1745-1750 (1989).
10. Schröder, J.-M., Sticherling, M., Henneicke, H. H., Preissner, W. C. & Christophers, E. *J. Immunol.* **144**, 2223-2232 (1990).
11. Akbar, A. N., Terry, L., Timms, A., Beverley, P. C. L. & Janossy, G. *J. Immunol.* **140**, 2171-2178 (1988).
12. Merckenschlager, M., Terry, L., Edwards, R. & Beverley, P. C. L. *Eur. J. Immunol.* **18**, 1653-1661 (1988).
13. Sanders, M. E. *et al.* *J. Immunol.* **140**, 1401-1407 (1988).

14. Cerottini, J.-C. & MacDonald, H. R. *A. Rev. Immun.* **7**, 77-89 (1989).
15. Graham, F. L., Smiley, J., Russell, W. L., Nairn, R. *J. gen. Virol.* **36**, 59-77 (1977).
16. Gorman, C. in *DNA Cloning Vol. II* (eds Glover, D.) 143-190 (IRL, Oxford, 1985).
17. Bacon, K. B., Camp, R. D. R., Cunningham, F. M. & Woollard, P. M. *Br. J. Pharmac.* **95**, 966-974 (1988).
18. Cordell, J. L. *et al.* *J. Histochem. Cytochem.* **132**, 219-229 (1984).
19. Sanders, M. E., Makgoba, M. W. & Shaw, S. *Immun. Today* **9**, 195-198 (1988).
20. Sanders, M. E., Makgoba, M. W., June, C. H., Yong, H. A. & Shaw, S. *Eur. J. Immunol.* **19**, 803-808 (1989).
21. Budd, R. C., Cerottini, J.-C. & MacDonald, H. R. *J. Immunol.* **138**, 3583-3586 (1987).
22. MacKay, C. R., Marston, W. L. & Dudler, L. *J. exp. Med.* **171**, 801-817 (1990).
23. Schagger, H. & Von Jagow, G. *Analyt. Biochem.* **166**, 368-379 (1987).
24. Unemori, E. & Werb, Z. *J. biol. Chem.* **263**, 16252-16259 (1988).

ACKNOWLEDGEMENTS. We thank E. Berry and S. Cuisia for preparing the manuscript, the Genentech peptide synthesis group for the RAN01 and RAN02 peptides, K. Andow and L. Tamayo for preparing the figures, and P. C. Beverley and L. Poulter for antibodies.

Intrinsic segmental identity of segmental founder cells of the leech embryo

Mark Q. Martindale* & Marty Shankland

Department of Anatomy and Cellular Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

* Present address: Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois 60637, USA

SEGMENTATION occurs in several animal phyla, and the cellular mechanisms generating this structural periodicity vary considerably¹⁻⁴. In the leech, an annelid worm, segmental founder cells arise through a fixed cell lineage (Fig. 1), and come together in a longitudinally repeating array through a stereotyped pattern of morphogenesis⁵. In this paper we demonstrate that founder cells forced to differentiate in a foreign segmental environment give rise to their normal, segment-specific clones of neuronal descendants, even in segments in which those neuronal phenotypes would not normally be observed. These findings indicate that the individual founder cells possess segmental identity at or shortly after the time of their birth, and further suggest that such identities are established by a mechanism in which the parent stem cell 'counts' mitotic cycles.

The central nervous system (CNS) of the leech is divided into 32 segments. The four most anterior segments (S1-S4) are fused to form the suboesophageal ganglion (Fig. 2a), which is followed by the unfused ganglia of the abdomen (A1, A2, and so on). In this paper, we use a monoclonal antibody⁶ against small cardioactive peptide B (SCP) to stain two segmentally distributed neurons in the leech *Helobdella*. The PLS1 neurons⁷ are bilaterally immunoreactive cells located in the posterolateral corner of each segmental ganglion. At the embryonic stages considered here, the PLS1 neurons stain intensely in segment S4, but stain faintly or not at all in the abdominal ganglia (Fig. 2a). The RAS neurons⁷ are unpaired cells situated in the anterolateral corner of S4 and A1-A3, but are not evident in more posterior segments. The unpaired RAS neurons arise from initially paired embryonic homologues through a process of competition^{8,9}, and typically alternate from left to right in successive ganglia (Fig. 2a).

The cell lineages of the leech embryo have been analysed in detail using intracellularly injected lineage tracers⁵, and the lineal origin of these neurons is known. Each hemilateral segment is generated by a small group of segmental founder cells—known as blast cells—which are themselves the daughters of a

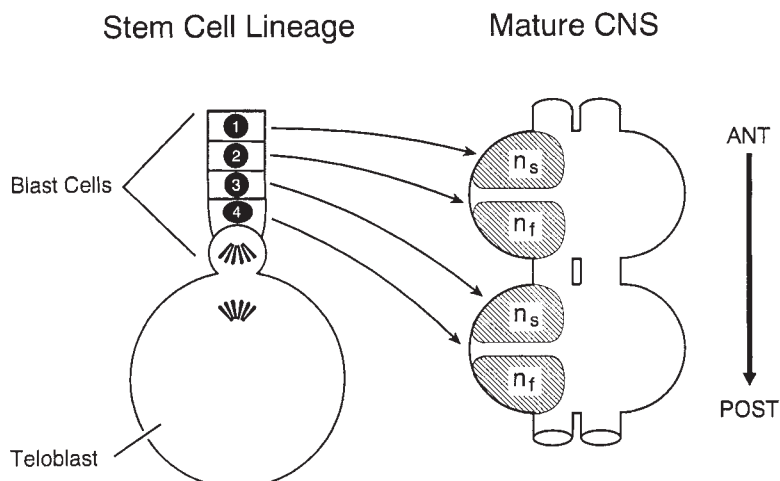
set of five large stem cells known as teloblasts. Figure 1 depicts the cell lineage of the N teloblast, which generates two blast cells per segment. One blast cell (n_s) contributes a clone of 70 descendant neurons to the anterior quadrant of a segmental ganglion, while the next blast cell (n_r) contributes a similarly sized clone to the posterior quadrant^{5,10}. The N teloblast generates n_s and n_r blast cells for successive segmental ganglia in anteroposterior order, and during normal development a blast cell's birth rank correlates precisely with the segmental location of its descendant clone. The PLS1 and RAS neurons are both descended from the N teloblast⁷, and their segment-specific patterns of immunoreactivity imply that neuropeptide expression is restricted to the descendants of certain specific n blast cells.

This segmental specificity could be explained if blast cells of differing birth rank have an intrinsic commitment to produce neurons of dissimilar phenotype. To test this possibility, we used the technique of 'bandlet slippage' to effectively transplant the n blast cells on one side of the embryo out of register with other segmental tissues¹¹, and then examined the distribution of immunoreactive neurons once the CNS had formed. Slippage was induced by photolesioning those n blast cells that contribute to the most anterior head segments, a perturbation which often results in the trailing n blast cells being frameshifted *en masse* so that they take part in the formation of abnormally posterior segments. The position of other segmental lineages—including the contralateral N lineage—is unaffected by this operation. Slippage was previously used to study segment-specific patterns of cell death, and the survival or death of particular cells was found to be specified by the host segment irrespective of lineage^{11,12}. In contrast, the present study focuses on segmental differences in neuronal differentiation, and the results discussed below indicate that segmental environment does not govern this aspect of blast cell fate.

The right or left N lineage was slipped at least one segment posteriorly in a total of 80 embryos. The PLS1 and RAS neurons were normally distributed on the unoperated side of the CNS, but on the operated side we routinely observed immunoreactive cell bodies located posterior to their normal domains (Figs 2b, d and 3a). Thus, blast cells that have been transplanted into foreign segments can give rise to neuronal descendants whose differentiated phenotypes are inappropriate for the host segment. Slippage of the n blast cells had little or no effect on the distribution of SCP-like immunoreactive neurons derived from other teloblasts, suggesting that different cell lineages establish segmental identity independently.

These changes in neuronal distribution were a direct result of blast cell slippage, not of the cell ablations that produced that slippage. For each embryo, the amount of slippage was

FIG. 1 The segmental tissues of the leech are generated by a small set of embryonic stem cells called teloblasts. Each teloblast undergoes an iterated sequence of asymmetric divisions to produce a chain of blast cell daughters which serve as the segmental founder cells for that lineage. During normal development, a blast cell's birth rank correlates precisely with the segmental location of its descendant clone. The N teloblast shown here generates two kinds of blast cell (n_s and n_r) which contribute clones to anterior and posterior quadrants, respectively, of each segmental ganglion in the mature CNS. The remainder of the ganglion arises from other stem cell lineages, including a symmetric contribution from the contralateral N teloblast.



measured with an accuracy of $\pm \frac{1}{2}$ segment by previously described techniques¹¹ (Fig. 2c legend). Of 17 operated embryos in which measured slippage was $\leq \frac{1}{2}$ segment, there were 14 cases in which the PLS1 and RAS distributions were normal and presumably no slippage had occurred, and 3 cases (assumed to be actual one-segment slips) in which these neurons were both displaced by only 1 segment. In contrast, of the 80 embryos in which measured slippage was ≥ 1 segment, all displayed ectopically immunoreactive neurons, and there were 42 cases in which such neurons were observed two or more segments posterior to their normal domains.

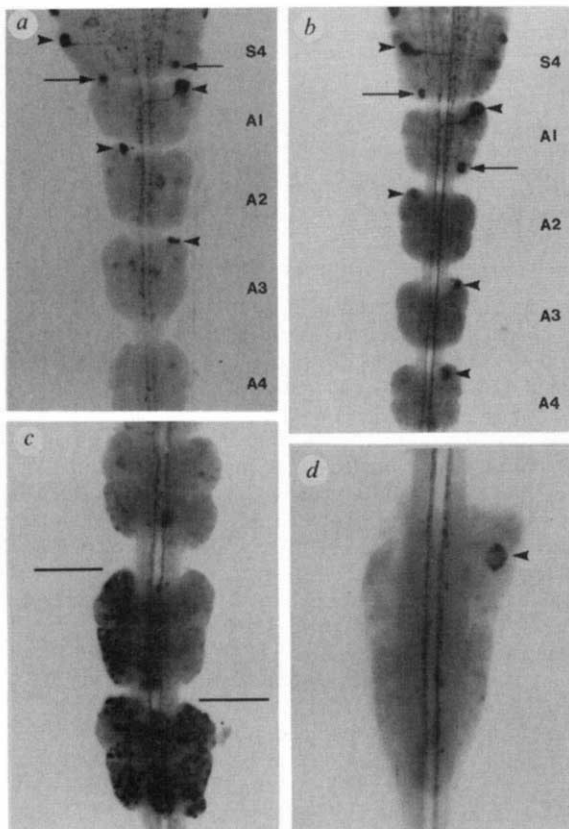


FIG. 2 Distribution of SCP-like immunoreactive cells in the CNS of normal and operated leech embryos. Anterior is to the top. *a*, In a normal embryo, anti-SCP intensely stains a bilateral pair of PLS1 neurons in segment S4 (arrows), with little staining of homologues in more posterior ganglia. In addition, a single RAS neuron (arrowhead) is stained in each of the ganglia S4–A3, but not in more posterior ganglia such as A4. *b*, In this embryo, the N cell lineage on the right side was slipped posteriorly with respect to the other segmental tissues, and there is a one-segment shift in the distribution of immunoreactive neurons (which arise from the slipped cell lineage) on the operated side. *c*, Same embryo as *b*, showing a one-segment disparity in the distribution of lineage tracer injected into the left and right N teloblasts. This disparity indicates that the right N lineage had slipped out of register posteriorly by $1 \pm \frac{1}{2}$ segment. *d*, Immunoreactive RAS neuron located in ganglion A7 of an embryo in which the right N lineage had been slipped four segments posteriorly.

METHODS. The CNS was dissected from formaldehyde-fixed embryos and stained with anti-SCP using fluorescein- or rhodamine-conjugated secondary antibodies⁷. Images presented are photographic negatives of fluorescent staining. In operated embryos, slippage of the right N lineage was accomplished by focal photolesions¹¹ of the most anterior n blast cells, which would normally have contributed to segments S1–S2. To measure slippage¹¹, the right and left N teloblasts were simultaneously injected with a rhodamine-dextran lineage tracer, and the anterior boundary of their labelled descendants compared, as in *c*. In unoperated embryos, the lineage tracer boundary is usually symmetric, and always falls within $\pm \frac{1}{2}$ segment (that is, the width of one blast cell clone) on the two sides. Disparities of ≥ 1 segment are indicative of slippage.

The experimentally induced distribution of immunoreactive neurons correlates well with the segmental location predicted by slippage alone (Fig. 3b). This relationship indicates that segmental identity is intrinsic to the slipped blast cell chain, and probably resides in the individual blast cells. The alternative would be that segmental identity is established by anteroposterior interactions along the blast cell chain, but this seems unlikely, given that we have never observed alterations of segmental specificity except with ablations that induce slippage. It should be noted that the segmental identity of the ancestral blast cell need not be passed along to PLS1 or RAS by a direct line of descent, as these neurons could depend for their segment-specific differentiation upon interaction with other derivatives of the slipped n blast cells which are linearly determined. In either case, we conclude that some segmental differences are determined by cell lineage, and that these primary differences most probably provide the template for other segmental specializations that involve interaction between cells of differing lineage^{12–14}.

Our results provide the first direct evidence that leech blast cells have an intrinsic segmental identity. Slippage was induced hours before the blast cells undergo their first cell division, and days before they would have come into register with other segmental tissues⁵. Bissen and Weisblat have suggested that the n_s and n_f blast cells are differentially specified as they are generated¹⁰, and it is now clear that segmental identity is also established at cell birth or shortly thereafter. If birth rank is the determinant of segmental identity in sequentially generated blast cells, it could imply that the underlying determinative mechanism counts the cell cycles of the parent stem cell and dictates

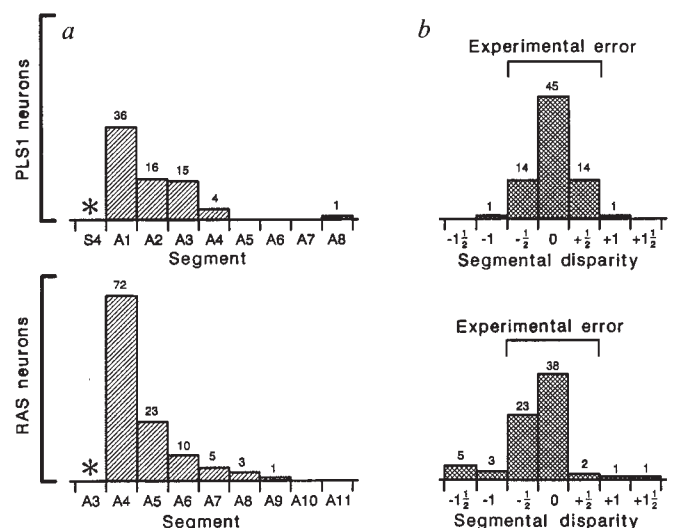


FIG. 3 Histograms showing the distribution of intensely immunoreactive PLS1 and RAS neurons on the operated side of leech embryos subjected to unilateral slippage of the N cell lineage. Numbers represent the total observations from a sample of 80 experimental embryos in which measured slippage was ≥ 1 segment. *a*, Segmental distribution of ectopic neurons. Asterisks mark the most posterior segment that displays an intensely immunoreactive neuron in the normal distribution. *b*, The location of ectopic neurons correlates closely with the measured slippage, suggesting that the slipped blast cells retain segmental identity and pass it along to their neuronal descendants. Brackets indicate the experimental error inherent to slippage measurements.

METHODS. For each experimental embryo, segmental disparity was calculated by comparing the observed location of the most posterior immunoreactive PLS1 and RAS neuron and the location predicted by the measured slippage. Positive disparity indicates that an immunoreactive neuron was observed posterior to the predicted segment; negative disparity indicates that the most posterior immunoreactive neuron was observed anterior to the predicted segment.

the identity of its daughters accordingly. A similar cell lineage mechanism has been invoked to explain the formation of birth rank-specified progeny by the neuroblast stem cells of insects¹⁵.

Specification of segmental identity is best understood in the *Drosophila* embryo, where individual segments become different through the regional activation of bithorax and antennapedia complex genes¹⁶. A homologous gene has been identified in the leech, and shown to have a segment-specific pattern of expression¹⁷, but it is not yet clear what functional role such genes have in leech segmentation. Our present findings indicate

that the segmental identity of leech blast cells is initially established through a process of temporal—rather than regional—specification, with this sequence of blast cell identities being translated into the anteroposterior sequence of segments by maintenance of birth rank along the blast cell chain. We would therefore predict that gene products that have a primary role in the specification of segmental identity are very probably expressed in the newly born blast cell and/or the parent teloblast, and that the temporal control of their expression may have a pivotal role in the differential specification of blast cells that are generated in succession. □

Received 19 April; accepted 20 August 1990.

1. French, V. *Development* **104** Suppl., 3–16 (1988).
2. Kimmel, C. B., Sepich, D. S. & Trevarrow, B. *Development* **104** Suppl., 197–208 (1988).
3. Stern, C. D., Fraser, S. E., Keynes, R. J. & Primm, D. R. N. *Development* **104** Suppl., 231–244 (1988).
4. Patel, N. *et al.* *Cell* **58**, 955–968 (1989).
5. Weisblat, D. A. & Shankland, M. *Phil. Trans. R. Soc.* **312**, 39–56 (1985).
6. Masinovsky, B., Kempf, S. C., Callaway, J. C. & Willows, A. O. D. *J. comp. Neurol.* **273**, 500–512 (1988).
7. Shankland, M. & Martindale, M. Q. *Dev. Biol.* **135**, 431–448 (1989).
8. Martindale, M. Q. & Shankland, M. *Dev. Biol.* **123**, 85–96 (1990).
9. Blair, S. S., Martindale, M. Q. & Shankland, M. *J. Neurosci.* **10**, 3183–3193 (1990).

10. Bissen, S. T. & Weisblat, D. A. *J. Neurobiol.* **18**, 251–269 (1987).
11. Shankland, M. *Nature* **307**, 541–543 (1984).
12. Martindale, M. Q. & Shankland, M. *Dev. Biol.* **125**, 290–300 (1988).
13. Loer, C. M., Jetties, J. & Kristan, W. B. *J. Neurosci.* **7**, 2630–2638 (1987).
14. Baptista, C. A. & Macagno, E. R. *J. Neurobiol.* **19**, 707–726 (1988).
15. Doe, C. Q. & Goodman, C. S. *Dev. Biol.* **111**, 206–219 (1985).
16. Akam, M. *Development* **101**, 1–22 (1987).
17. Wysocka-Diller, J. W., Aisemberg, G. O., Baumgarten, M., Levine, M. & Macagno, E. R. *Nature* **341**, 760–763 (1989).

ACKNOWLEDGEMENTS. This work was supported by the NIH and March of Dimes (M.S.); M.Q.M. was supported by an NIH postdoctoral fellowship.

Cloning of a gene that is rearranged in patients with choroideraemia

Frans P. M. Cremers, Dorien J. R. van de Pol, Liesbeth P. M. van Kerkhoff, Berend Wieringa & Hans-Hilger Ropers

Department of Human Genetics, University Hospital, University of Nijmegen, PO Box 9101, 6500HB Nijmegen, The Netherlands

CHOROIDERAEMIA (tapetochoroidal dystrophy, TCD), a common form of X-linked blindness¹, is characterized by progressive dystrophy of the choroid, retinal pigment epithelium and retina^{2,3}. Previous studies have assigned the TCD gene to a small segment of the Xq21 band^{4–6}. By making use of reverse genetics strategies we have isolated eight overlapping complementary DNA clones from the same chromosomal region. The corresponding gene is expressed in retina, choroid and retinal pigment epithelium. The cDNAs encompass an open reading frame of 948 base pairs that is structurally altered in eight TCD patients with deletions, and in a female patient with a balanced translocation involving Xq21. These findings provide strong evidence that we have cloned the gene underlying choroideraemia. Elucidation of its function should provide new insights into the molecular mechanisms responsible for this disorder and other hereditary retinopathies.

In previous studies we and others have shown that the TCD gene maps to an interval of the Xq21 band defined by the DNA markers DXS95, DXS165 and DXS233^{7,8}. One of these loci, DXS165, is deleted in several patients with classical TCD^{4–6}. As a prerequisite for the molecular characterization of the TCD gene region, deletion-breakpoint cloning⁹, preparative field inversion gel electrophoresis¹⁰, and walking and jumping techniques^{5,11} were used to generate new DNA markers near the DXS165 locus. Using these new markers, a genomic segment of approximately 45 kilobases (kb) which overlaps most of the TCD-associated deletions (Fig. 1) could be cloned. From this segment, 15 single-copy sequences were isolated and screened for evolutionary conservation by hybridization to genomic DNA from various vertebrate species. With two of these DNAs, probes 398 and 413 (Fig. 1), specific hybridization signals were obtained with DNAs from several species including chicken (Fig. 3a). On northern blots, probe 398 yielded distinct signals in RNAs

from retina and retinal cell lines (not shown). Screening of a human retinal cDNA library¹² resulted in the isolation of eight overlapping cDNA clones (T1–T8), which span a total of 4.5 kb but do not contain the 5' and 3' ends of the gene (Fig. 1). Clone T8 extends furthest upstream and contains an open reading frame (ORF) (Fig. 2a, ORFI) capable of encoding a polypeptide of 316 amino acids. Nuclease S1 analysis indicated that the insert of this clone is fully protected by messenger RNA from a retinal cell line up to about 55 nucleotides from its 5' end (Fig. 2b). The sequence immediately upstream of position 56 (Fig. 2a; underlined) has a conspicuous resemblance to the consensus sequence for 3' splice sites¹³, which may indicate that these nucleotides originate from an intron extending further upstream. At position 61–63, there is a stop codon, and the first ATG codon of the ORF at position 127–129 may function as a translation start site¹⁴.

None of the cDNA clones contains the poly(A) sequences normally indicative of the 3' end of mRNAs. The complete sequence deduced from the combined cDNAs (not shown) contains a strikingly large 3' untranslated region of 3.4 kb. In the middle of this region, starting at position 2,800 and adjacent to a sequence that could act as a 3' splice site, we have found an ORF spanning 395 base pairs (bp), (Fig. 1; ORFII). The functional significance of this sequence remains to be elucidated.

To ascertain the relationship between the cloned cDNA and the TCD gene, genomic DNAs of eight patients with TCD who carried submicroscopic deletions of the Xq21 band^{5,6} were re-examined with single-copy subclones of cDNA T1. In all patients the deletions had removed at least part of the long ORF that spans segments A to E of the consensus cDNA (Figs 1 and 3b). In addition, we could locate the position of the X-chromosomal breakpoint of a t(X;13) translocation previously identified in a female with typical signs of TCD^{5,15}, to intron sequences that separate segments C and D of the cDNA (Fig. 3c). In combination, these data provide strong evidence that the cloned cDNA is indeed part of the choroideraemia gene.

RNA analyses have corroborated this conclusion. In choroid/retinal pigment epithelium, retina, two retinal cell lines (HER XC2 and HER RC2; ref. 16), and HeLa cells, all eight cDNA clones hybridize to an mRNA transcript of ~5,400 residues (Fig. 4a). Surprisingly, this mRNA is also expressed in Epstein-Barr virus-immortalized B cells, albeit at a significantly lower level. This finding enabled us to study mRNA patterns in patients with deletions of various size and should pave the way for the detection of point mutations in the TCD gene. As shown in Fig. 4b, with a probe from the 5' end of the

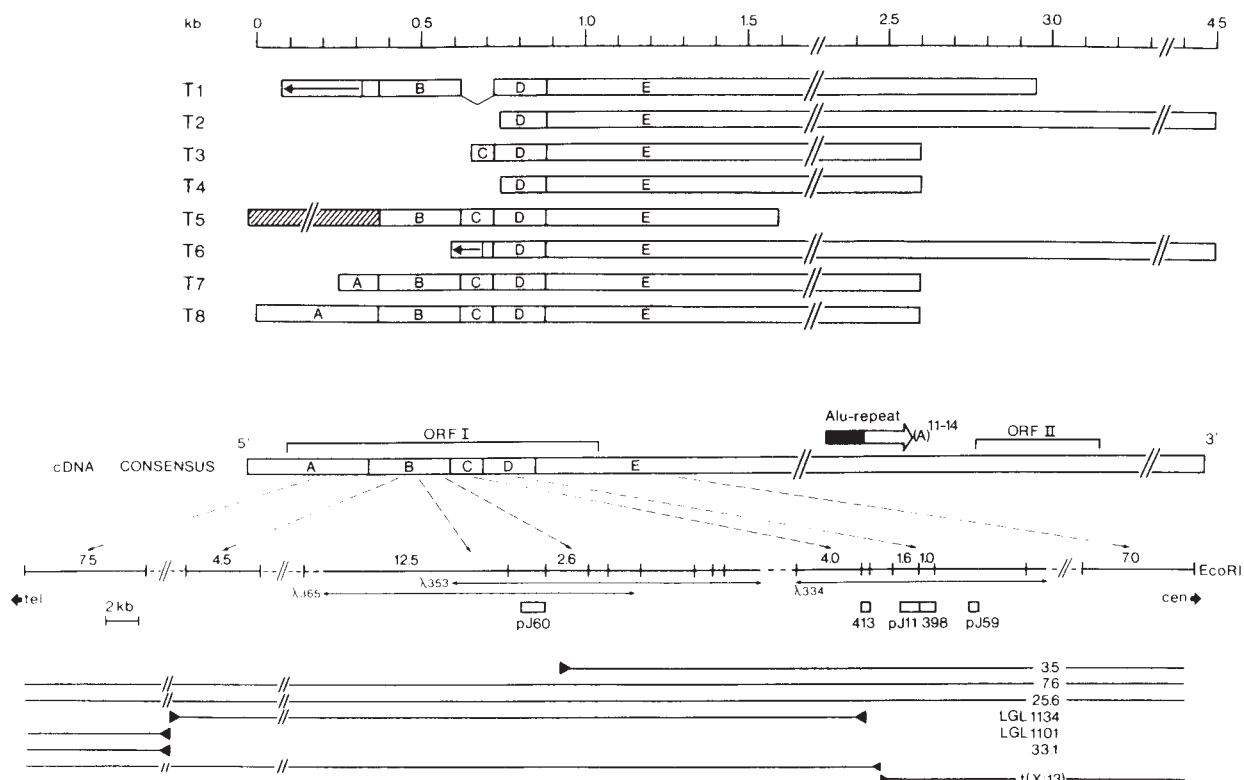


FIG. 1 Overlapping cDNA clones, their alignment with respect to genomic DNA fragments as well as deletion and translocation breakpoints in the Xq21 region. Top, the relative positions of the eight cDNAs shown were deduced from restriction enzyme mapping and sequence analysis. Genuine exon sequences in cDNA inserts are represented by open boxes. A-E denote distinct cDNA segments, each of which may be composed of several exons. Vertical bars represent exon-exon boundaries as inferred from comparison with genomic DNA sequences (clones pJ11 and 398 for defining 5' and 3' borders of segment D, respectively) or from comparisons between different cDNAs (T1, T5, T7, and T8 for both borders of segment C and the 3'-border of segment A). The cross-hatched box at the 5' end of T5 probably represents an intron sequence, as a perfect splice acceptor splice site sequence¹³ was identified directly upstream of segment B in this clone. Four clones (T3, T4, T7 and T8) with truncated 3' ends probably originate from oligo(dT) primed reverse transcription starting at a poly(A) stretch located just 3' of an internal Alu repeat, which is indicated by a boxed arrow. Clones T1 and T5 probably result from alternatively or aberrantly spliced precursor RNAs. Both the segment lacking from T1 (segment C) and the extra stretch identified at the 5' end of T5 (cross-hatched segment) feature genuine splice site signals¹³ at their borders. Arrows in clones T1 and T6 indicate sequences that are reversed with respect to the sequence determined for the combined cDNAs. These reversed segments contain short inverted repeats at their ends and probably originate from aberrant processing of stem-loop structures during reverse transcription of the mRNA. Bottom, the consensus cDNA as deduced from eight cDNA clones is shown in relation to a partial genomic map of the corresponding chromosome segments. Deletions are shown as horizontal bars between triangles. The X-chromosomal breakpoint in a female with a balanced t(X;13) translocation^{5,15} is indicated. The positions of the two longest ORFs in the cDNA are given. The partial *EcoRI*

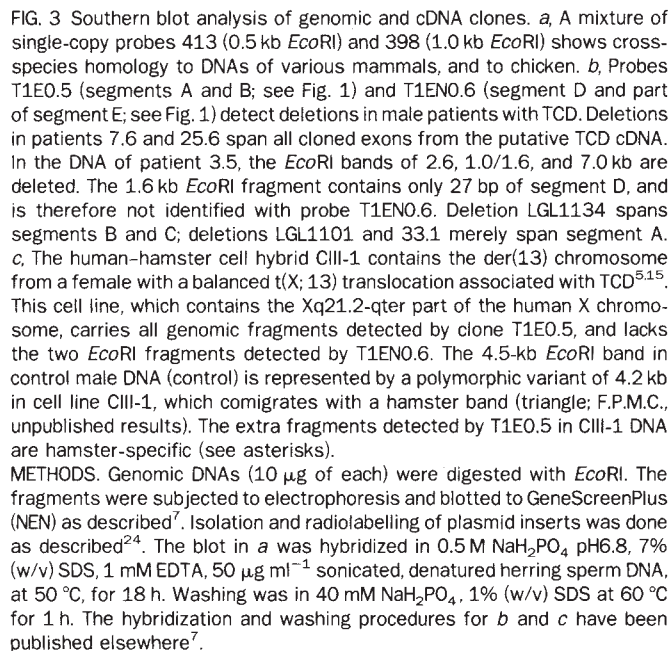
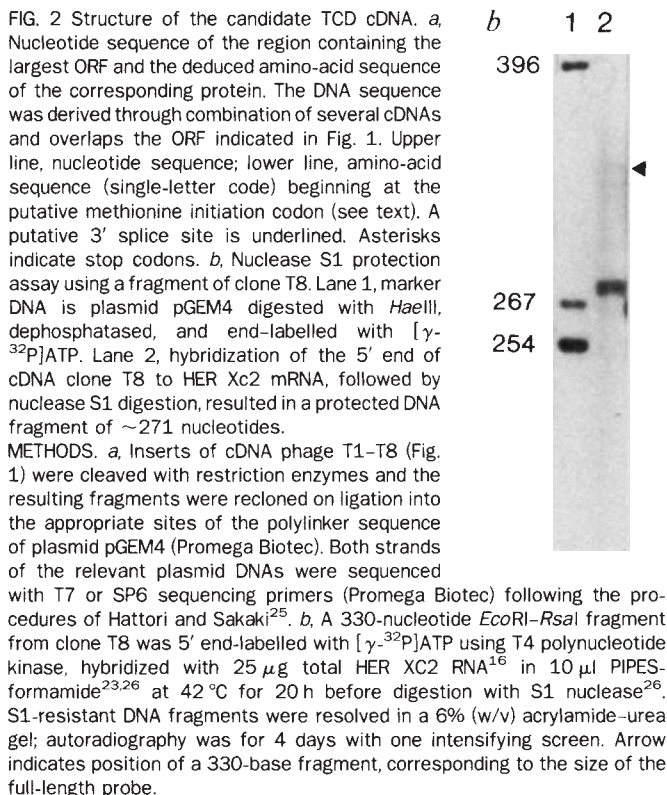
map was constructed by Southern analysis of *EcoRI*-digested genomic DNA, and by restriction analysis of three recombinant phages (λ 334, λ 353 and λ 365). The chromosomal origin as well as the telomere-centromere orientation of the cDNA segments A-E is indicated. The outermost sections of the cDNA are homologous to *EcoRI* fragments that are still unlinked to the central region of the genomic map.

METHODS. Genomic DNA phage clones with human inserts were isolated employing jump clones pJ60, pJ11 and pJ59 (see ref. 5) using standard cloning methodology. Phage 353 and 365 originated from a total human DNA genomic library in phage EMBL3 (courtesy of G. Grosveld, Rotterdam). Phage 334 was isolated from a pool of recombinant λ DASH (Stratagene) phage containing 14.5 kb *HindIII* inserts enriched by preparative gel electrophoresis (for methods, see ref. 9). Plaque purification, DNA isolation, restriction enzyme mapping and subcloning of phage-insert fragments into pGEM 3/4 plasmid vectors (Promega Biotec) was done according to standard recombinant DNA procedures²³. Single-copy *EcoRI* inserts identified in plasmids 398 and 413 were subsequently used as probes on zoo blots (see Fig. 3a) and for screening a λ gt10 human retinal cDNA library¹². Phage recombinants (10⁶) were screened with a mixture of ³²P-labelled²⁴ 398 and 413 insert DNAs using standard procedures²³. Eight positive cDNA clones, T1-T8, were plaque-purified. Insert DNA prepared from plate lysates of each of the eight recombinant phage was cleaved with *EcoRI* and subcloned in plasmid pGEM4. Alignment of various cDNAs was based on restriction mapping using restriction enzymes *AccI*, *BglII*, *EcoRI*, *HindIII*, *NcoI* and *PstI* and on sequence analysis determination (see Fig. 2). The region between nucleotide positions 1 and 1,000 shown at the top of the figure and a stretch of ~250 nucleotides at the 3' ends were sequenced in all cDNA clones. Only for clone T6 was the insert completely sequenced.

cDNA (T1E0.5), no hybridization signals were seen on northern blots containing RNAs from lymphoblastoid cell lines of TCD patients 7.6 and 25.6. This is not unexpected as both deletions span the entire gene. In contrast, patient 3.5, who lacks the 3' end of the gene, shows aberrant transcripts of about 4.5 and 5.5 kb. In this patient, two different novel sites may be used as polyadenylation signals.

Neither the nucleotide nor the predicted amino-acid sequence of the putative gene product revealed any significant homology to genes or protein sequences in the NBRF (December 1989), Swiss-Prot (January 1990) and EMBL (April 1990) databases.

The protein carries potential phosphorylation sites for protein kinase C (residues 76, 122, 175, 178, 301) and for casein kinase (residues 13, 129, 178, 195, 239, 301), but no other sites associated with post-translational modification processes. Moreover, no topogenic sequences, no domains with known biological function and no enzymatic active sites could be determined using the PCGene PROSITE program. The C-terminal portion of the protein between residues 253 and 302 carries a so-called PEST region (PEST score of 7.3 according to Rogers *et al.*¹⁷). Proteins containing one or more PEST regions often exhibit intracellular half lives of less than 2 h, and several of these, such



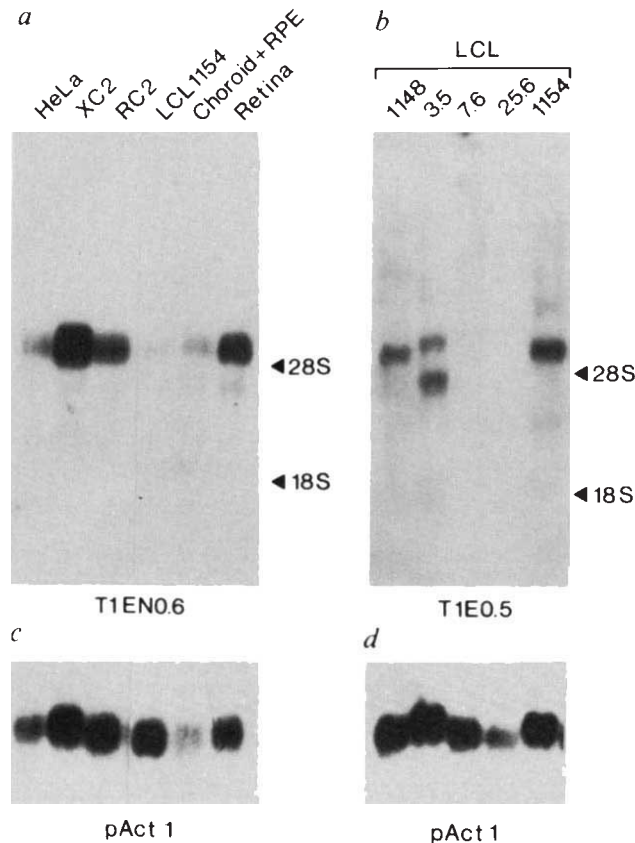


FIG. 4 Northern blot analysis of RNA from several human cell lines and tissues using cDNA clone T1. *a*, Blot containing RNA from HeLa, two retinal cell lines (HER RC2 and HER XC2; ref. 16), an EBV-immortalized lymphoblastoid cell line (LCL1154), as well as human choroid/retinal pigment epithelium and retina. *b*, Blot containing RNA from lymphoblastoid cell lines from patients with different microdeletions encompassing the TCD locus (Fig. 1, and refs 5 and 6), and from two male controls (LCL1148 and LCL1154). *c*, *d*, Hybridization of a hamster actin cDNA clone (pAct-1; ref. 27) to the same blots as in *a* and *b*, respectively. The positions of the 18S and 28S ribosomal RNA bands are indicated.

METHODS. Total cellular RNA was isolated from cells and tissues using the modified LiCl-urea extraction procedure^{28,29}. Approximately 10 µg of RNA was dissolved in 30 µl of 50% (v/v) formamide, 2.2 M formaldehyde, 20 mM MOPS, pH 7.0 buffer (4-morpholine-propanesulphonic acid, 5 mM Na-acetate, 1 mM EDTA), loaded onto a 1% agarose gel containing 2.2 M formaldehyde in 20 mM MOPS buffer, and resolved by electrophoresis at 1 V cm⁻¹ for 16 h. The gel was rinsed extensively with sterile water for 15 min and RNA blotted to a GeneScreenPlus membrane (NEN) in 10 × SSC. After baking for 2 h at 80 °C, blots were hybridized with random-primed probes. Prehybridization and hybridization was done in 5 × SSPE, 10% (w/v) dextran sulphate, 1% (w/v) SDS, and 100 µg ml⁻¹ sonicated, denatured herring sperm DNA at 60 °C for 4 and 16 h respectively. Washing was done at the same temperature in 2 × SSC, 1% (w/v) SDS for 40 min.

as Fos, Myc and E1A, are important transcriptional regulators^{17,18}.

Recently, photoreceptor-specific genes have been implicated in autosomal dominant retinitis pigmentosa¹⁹ and two different forms of retinal degeneration of the mouse^{20,21}. It is of note, therefore, that the expression of the putative TCD gene is not confined to the eye. This finding is not unprecedented, however. For example, gyrate atrophy, a choroidal disease with clinical similarity to TCD, is caused by a deficiency of an ubiquitously expressed protein, the enzyme ornithine aminotransferase²². By analogy, this may indicate that TCD also is due to a generalized metabolic defect. As this is the first candidate gene for any of the X-linked forms of chorioretinal degeneration, elucidation of its structure and function may provide deeper insight into the molecular mechanisms underlying degenerative processes of the retina, retinal pigment epithelium and choroid. □

Received 4 July; accepted 30 August 1990.

- Heckenlively, J. R. *Retinitis pigmentosa* (Lippincott, Philadelphia, 1988).
- Goedbloed, J. *Ophthalmologica* **104**, 308–315 (1942).
- Waardenburg, P. J. *Acta ophthalmol.* **20**, 235–274 (1942).
- Cremers, F. P. M. et al. *Clin. Genet.* **32**, 421–423 (1987).
- Cremers, F. P. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7510–7514 (1989).
- Cremers, F. P. M. et al. *Am. J. hum. Genet.* (in the press).
- Cremers, F. P. M. et al. *Genomics* **4**, 41–46 (1989).
- Merry, D. E. et al. *Am. J. hum. Genet.* **45**, 530–540 (1989).
- Cremers, F. P. M. et al. *Hum. Genet.* (in the press).
- van de Pol, T. J. R., Cremers, F. P. M., Brohet, R. M., Wieringa, B. & Ropers, H.-H. *Nucleic Acids Res.* **18**, 725–731 (1990).
- Merry, D. E. et al. *Genomics* **6**, 609–615 (1990).
- Nathans, J., Thomas, D. & Hogness, D. S. *Science* **232**, 193–202 (1986).
- Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. *Rev. Biochem.* **55**, 1119–1150 (1986).
- Kozak, M. J. *molec. Biol.* **196**, 947–950 (1987).
- Siu, V. M., Gonder, J. R., Jung, J. H., Sergovich, F. R. & Flintoff, W. F. *Hum. Genet.* **84**, 459–464 (1990).
- Vaessen, R. T. M. J., Houweling, A., Israel, A., Kourilsky, P. & van der Eb, A. J. *EMBO J.* **5**, 335–341 (1986).
- Rogers, S., Wells, R. & Rechsteiner, M. *Science* **234**, 364–368 (1986).
- Rechsteiner, M., Rogers, S. & Rote, K. *Trends Biochem. Sci.* **12**, 390–394 (1987).
- Dryja, T. P. et al. *Nature* **343**, 364–366 (1990).
- Travis, G. H., Brennan, M. B., Danielson, P. E., Kozak, C. A. & Sutcliffe, J. G. *Nature* **338**, 70–73 (1989).
- Bowes, C., Danciger, M., Kozak, C. A. & Farber, D. B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9722–9726 (1989).
- Valle, D. & Simell, O. In *The Metabolic Basis of Inherited Disease* (eds Stanbury, J. B., Wjingaarden, J. B., Fredrickson, D. S., Goldstein, J. L. & Brown, M. S.) 382–401 (McGraw-Hill, New York, 1983).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **137**, 266–267 (1984).
- Hattori, M. & Sakaki, Y. *Analyt. Biochem.* **152**, 232–238 (1986).
- Weaver, R. F. & Weissman, C. *Nucleic Acids Res.* **7**, 1175–1193 (1979).
- Dodemont, H. J. et al. *EMBO J.* **1**, 167–171 (1982).
- Auffray, C. & Rougeon, F. *Eur. J. Biochem.* **107**, 303–314 (1980).
- LeMeur, M. et al. *Cell* **23**, 561–571 (1981).

ACKNOWLEDGEMENTS. We thank J. Nathans for the cDNA library, F. S. Collins for the jumping library, A. van der Eb for the retinal cell lines HER XC2 and HER RC2, G. Grosfeld for the genomic library, E.-M. Sankila and F. Brunsmann for DNA of patients with TCD, and F. Oerlemans and L. van Rossum for technical assistance. We also thank L. Pels for eye tissues, J. G. W. Domen for running protein analysis programs, and P. N. Goodfellow for a critical reading of the manuscript. The help of J. A. M. Leunissen and the use of the services and facilities of the Dutch CAOS/CAMM centre is gratefully acknowledged. This work was supported by the Nederlandse organisatie voor Wetenschappelijk Onderzoek, Landelijke Stichting voor Blinden en Slechtzienden, Gelderse Blinden Vereniging, Blinden-Penning and the British Retinitis Pigmentosa Society.

Retinal degeneration in the *rd* mouse is caused by a defect in the β subunit of rod cGMP-phosphodiesterase

Cathy Bowes*, Tiansen Li†, Michael Danciger‡*, Leslie C. Baxter†, Meredith L. Applebury† & Debora B. Farber*§

* Jules Stein Eye Institute, University of California, Los Angeles, School of Medicine, Los Angeles, California 90024, USA

† The University of Chicago, Eye Research Laboratories, 939 East 57th Street, Chicago, Illinois 60637, USA

‡ Loyola Marymount University, Los Angeles, California 90045, USA

MICE homozygous for the *rd* mutation display hereditary retinal degeneration and the classic *rd* lines serve as a model for human retinitis pigmentosa^{1,2}. In affected animals the retinal rod photoreceptor cells begin degenerating at about postnatal day 8, and by four weeks no photoreceptors are left³. Degeneration is preceded by accumulation of cyclic GMP in the retina⁴ and is correlated with deficient activity of the rod photoreceptor cGMP-phosphodiesterase⁵. We have recently isolated a candidate complementary DNA for the *rd* gene⁶ from a mouse retinal library and completed the characterization of cDNAs encoding all subunits of bovine photoreceptor phosphodiesterase⁷. The candidate cDNA shows strong homology with a cDNA encoding the bovine phosphodiesterase β subunit. Here we present evidence that the candidate cDNA is the murine homologue of bovine phosphodiesterase β cDNA. We conclude that the mouse *rd* locus encodes the rod photoreceptor cGMP-phosphodiesterase β subunit.

§ To whom correspondence should be addressed.

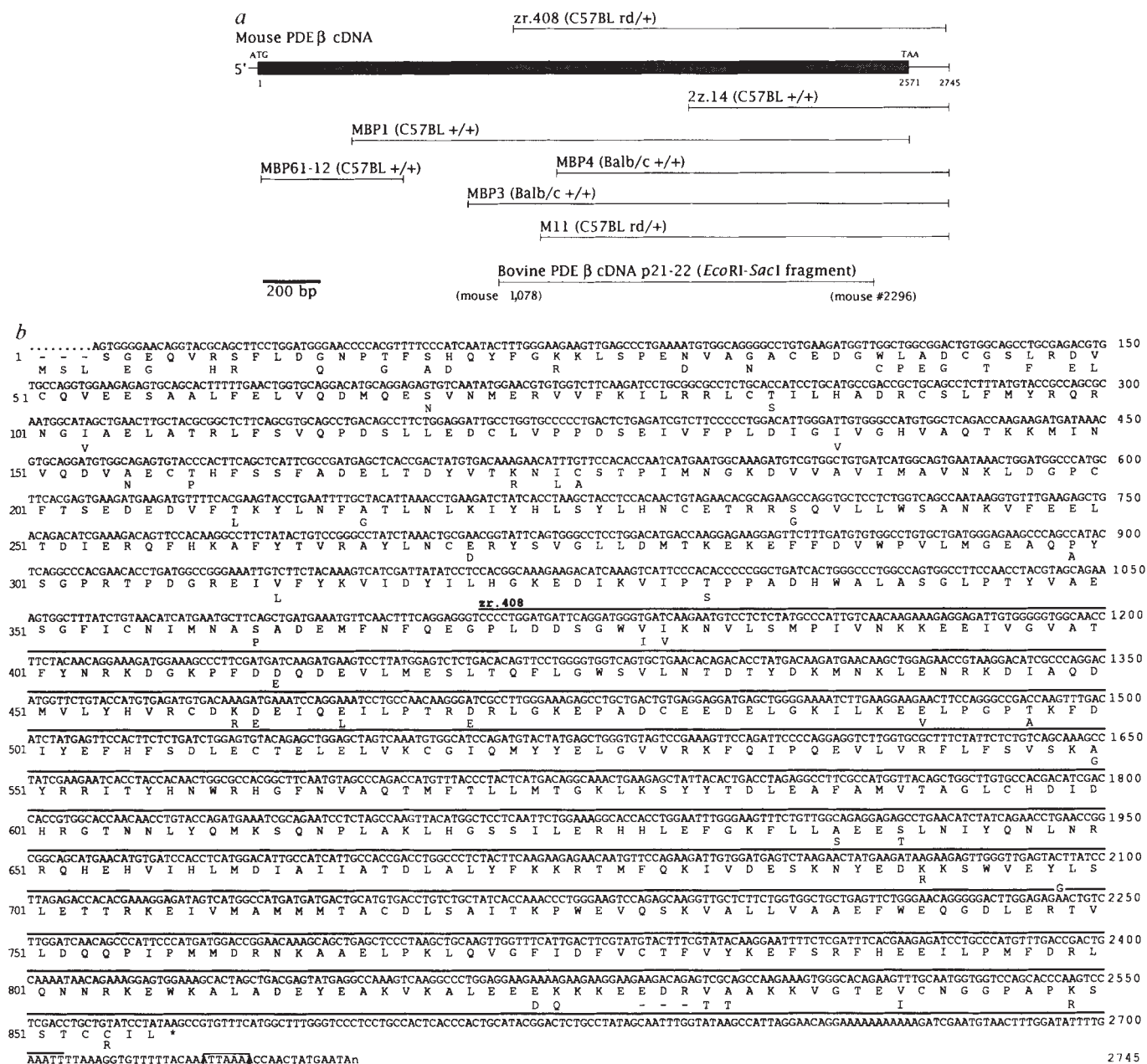


FIG. 1. Structure of the mouse photoreceptor cGMP-PDE β subunit cDNA and comparison with the *rd* candidate cDNA, zr.408, a, Diagram of the PDE β cDNA and positions of cDNA clones relevant to this work. The 2,569-base open reading frame is represented by the filled box. The mouse strains from which the cloned inserts were derived are given in parentheses. The fragments are positioned with respect to the PDE β cDNA structure based on their homology alignment; zr.408 represents the '1.65-kb' *EcoRI* fragment described as the *rd* candidate cDNA⁶. The nucleotide sequence of the mouse PDE β subunit is compiled from MBP1, MPB61-12, and M11. The probes zr.408, 2z.14, MBP3, MBP4, and the *EcoRI*-*SacI* fragment of bovine PDE β p21-22 clone¹⁷ were used for Southern (Fig. 2) and northern (Fig. 3) blots. M11 and zr.408, derived from heterozygous *rd*/+ libraries, can be distinguished as wild type (+) on the basis of a minor polymorphism in the 3' untranslated end (unpublished data). b, Nucleotide sequence and translated amino-acid sequence (single-letter notation) of mouse cGMP-PDE β subunit cDNA. The sequence is numbered from the translation initiation codon ATG on the basis of the bovine homologue⁷ position. The amino-acid sequence differences in the bovine PDE β -subunit are given below the mouse sequence. The line above the nucleotide sequence represents zr.408. A common variant of the polyadenylation signal, ATTAAG, at 19 bases before the polyadenylation site, is boxed.

METHODS. Isolation of zr.408 has been described⁶; a 1.6-kb *EcoRI* restriction fragment was subcloned into pBluescript SK(-). Complementary DNA libraries

were constructed in λ Zap cloning vectors and pBluescript SK(-) phagemids with inserts were excised *in vivo* from plaque-purified λ Zap phage clones according to supplier's instructions (Stratagene). 2z.14 was isolated from a 28-day-old C57BL/6J+/+ mouse retina cDNA library in λ Zapl using the zr.408 probe with hybridization conditions of 7.0% SDS, 0.5M phosphate buffer, pH 7.0, 1.0mM EDTA, 1% BSA at 65°C and washed to a final stringency of $0.2 \times \text{SSC}/0.1\%$ SDS at 60°C. MBP3 and MBP4 were isolated from a BALB/c+/+ retinal library in λ Zapl and M11 from a C57BLrd+/+ retinal cDNA library in λ Zapl using the bovine p21-22 probe. Libraries were screened at high stringency as described¹⁸; final washes were with $0.2 \times \text{SSC}/0.1\%$ SDS at 60°C. MBP1 and MBP61-12 were obtained by PCR of first-strand cDNA synthesized from C57BL/6J+/+ retinal mRNA carried out with a Perkin Elmer-Cetus GeneAmp kit and DNA thermal cycler. The 5' oligonucleotide primers were synthesized from the bovine PDE β sequences⁷ and 3' primers from mouse PDE β sequences (b). Because the bovine 5' primer extended into the coding region, identification of mouse sequences near the initiation site was precluded. Amplified DNA products were cloned into the pBluescript SK(-) *Sma*I site. Sequencing was by the dideoxy chain termination method¹⁹ using the Sequenase kit from US Biochemicals. Both strands of DNA were fully sequenced. Sequence data were analysed with the Genetics Computer Group Sequence Analysis Software, version 5.2 (ref. 20) or with the MICROGENIE Sequence Analysis Program (Beckman).

Cyclic GMP is the key messenger linking photon absorption to neural signalling in vertebrate photoreceptors. Absorption of a photon by rhodopsin triggers the sequential activation of transducin and cGMP-phosphodiesterase (PDE). Depletion of cGMP results in closure of cGMP-gated ionic channels, and thus in visual signalling⁸. In theory, a defect in any of the components in this pathway might cause abnormal cGMP metabolism such as that associated with the *rd* lesion⁴. Studies of rhodopsin, transducin and 48K protein (a protein of relative molecular mass 48,000, involved in rhodopsin deactivation) have shown that there are no differences in normal and *rd* mouse retinæ, as judged by the size and relative abundance of their messenger RNAs or presence of protein products^{9,10}. Studies of the cGMP-PDE in *rd* retinæ have shown, however, that the enzyme has low activity⁵, and is present in reduced concentration¹¹, making it a prime candidate for the *rd* mutation. The enzyme is a complex of two large catalytic subunits, α and β , and two copies of a small inhibitory subunit, γ (refs 12, 13). We have shown that the cGMP-PDE α and γ subunits map to mouse chromosomes 18 and 11, respectively^{14,15}. As the *rd* locus is on chromosome 5 (refs 1, 3), between the anchor genes *Afp* and *Gus*, α and γ cannot be candidates for the *rd* gene. The recent characterization of the β subunit⁷ made it possible to examine this subunit as a potential site of the defect.

A candidate *rd* cDNA, zr.408, maps to the same chromosomal location as the *rd* gene, hybridizes to a slightly larger, less abundant transcript in the developing *rd* retina compared with normal retina, and detects restriction fragment-length polymorphisms (RFLPs) in genomic DNA that segregate with the *rd* phenotype^{6,16}. Initial exploration for homologies between zr.408 and cDNAs for the bovine rod PDE α , β and the cone PDE α' subunits indicated strong cross-hybridization with β cDNA. To pursue this observation, zr.408, which was isolated from a heterozygous C57BL/6J *rd*/+ mouse retinal cDNA library, was used as a probe to obtain cDNA clones from an adult C57BL/6J+/+ retinal cDNA library. A bovine PDE β cDNA probe, p21-22, was used to identify murine homologous cDNA clones in BALB/c and C57BL/6J+/+ retinal cDNA libraries (Fig. 1a).

A nucleotide sequence and deduced amino-acid sequence were compiled from overlapping murine homologue cDNA clones identified with the bovine PDE β probe (Fig. 1). As previously described for the isolation of bovine cGMP-PDE β cDNA⁷, the many clones isolated were truncated and no full-length clones for the mouse homologue were identified. Polymerase chain reactions (PCR) were used to provide fragments covering the full mRNA length. The mouse sequence shows 93.1% identity with the bovine β polypeptide, as opposed to 73.3% identity with the PDE α polypeptide or 64% identity with bovine cone PDE polypeptide¹⁷ and confirms that the mouse clones represent the cGMP-PDE β transcript. The sequence of zr.408 encompasses bases 1,114 through to 2,706 of the mouse PDE β cDNA (Fig. 1b). It is identical except for one nucleotide, which may arise from a polymorphic difference in mouse strains used for library preparation. Thus, zr.408 is a truncated cDNA corresponding to the 3' two-thirds of the mouse PDE β transcript (Fig. 1b).

Genomic DNA from *rd* and normal murine tissues were probed in parallel with the bovine and murine PDE β and zr.408 DNA fragments for comparison. The four probes tested detect a common set of patterns in the genomic Southern blot, including the RFLPs that distinguish the *rd* and wild-type genotypes as shown for an *Msp*I restriction digest (Fig. 2). The 2.4-kilobase (kb) and 1.9-kb *Msp*I fragments are counterparts of the *rd* and wild-type genomes. They contain both introns and exons that map to the 3' end of the genes and encompass nucleotides 2,296 through to 2,743 in the PDE β cDNA (Fig. 1b) (unpublished data); the bovine probe fails to detect these fragments because it does not include these 3' sequences. The 1.4-kb fragment is detected by the zr.408 and bovine probes only because they

extend further 5' beyond probes MBP4 and 2z.14 (Figs 1a and 2). The origin of the 3.8-kb band that arises only with bovine probe is attributed to cross-hybridization with a related gene fragment, which occurs under the lower hybridization stringency conditions necessary to obtain good signals with the bovine probe.

Likewise, a mouse cGMP-PDE β probe and the zr.408 probe show similar patterns in northern blots of murine poly(A)⁺ RNA (Fig. 3). Both detect a broad mRNA band of ~3.1 kb in wild-type +/+ retinæ, but both show alteration in the abundance and size distribution of mRNA in the *rd*/*rd* (lanes 2, 5) and heterozygous *rd*/+ retinæ (lanes 3). The mRNA in the predegenerate *rd* retina is considerably reduced, as noted in Fig. 3, where less hybridization occurs with *rd*/*rd* than with wild-type mRNA, even though 5- or 10-fold more sample is loaded; likewise hybridization with the heterozygous *rd*/+ sample is at best equivalent to wild type, although fivefold more sample is loaded. The probes fail to detect any mRNA in adult *rd* retinæ (lanes 4) in which the photoreceptors are absent. In addition, the size

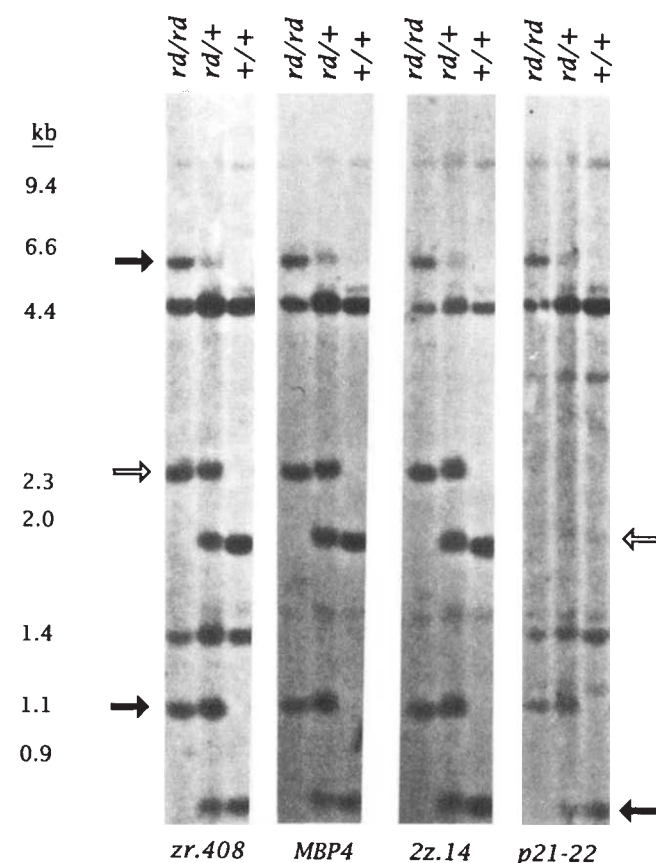


FIG. 2 Southern blots of mouse genomic DNA digested with *Msp*I and hybridized with zr.408, MBP4, 2z.14 or p21-22 (bovine PDE β clone) probes as shown in Fig. 1a. Lanes: *rd*/*rd*, *rd*/+, and +/+ indicate the C57BL/6J mouse strain used for DNA isolation. Arrows on the right mark polymorphisms common to +/+ and *rd*/+, but absent from *rd*/*rd*; arrows on the left mark polymorphisms common to *rd*/*rd* and *rd*/+, but absent from +/+. Filled arrows mark polymorphisms detected by all four probes; open arrows mark polymorphisms detected by all probes except p21-22. Standards for DNA mobility (on left, in kb) are a *Hind*III digestion of λ DNA and a *Hae*III digestion of ϕ X174 DNA.

METHODS. DNA was isolated and handled as previously described¹⁶, digested with *Msp*I, loaded at 6 μ g per lane in 1% agarose gels, transferred to nylon membranes and hybridized with random-primed labelled probes^{6,16}. Blots were hybridized as described for screening methods in Fig. 1, except that hybridization was carried out at 65 °C. They were pre-washed for 15 min at 37 °C in 2 $\times$ SSC/0.3% SDS. With zr.408, 2z.14 or MBP4 probes, blots were washed twice more for 20 min at 57 °C in 2 $\times$ SSC/0.3% SDS and twice for 20 min in 0.2 $\times$ SSC/0.3% SDS; with the bovine probe blots were also washed twice for 20 min at 47 °C in 2 $\times$ SSC/0.3% SDS, and twice for 20 min at 47 °C in 0.2 $\times$ SSC/0.3% SDS.

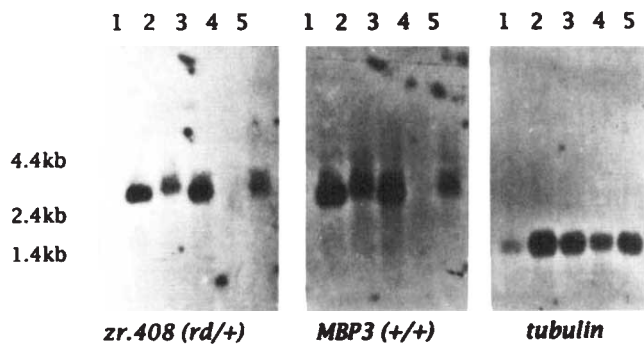


FIG. 3 Northern blot analysis of mouse poly(A)⁺ RNA probed with zr.408, MBP3 and human α -tubulin cDNA. (1) +/+ 28-day-old (0.2 μ g RNA); (2) *rd/rd* 10-day-old (predigested) (2 μ g); (3) *rd/+* 10-day-old (1 μ g); (4) *rd/rd* > 30-day-old (0.5 μ g); (5) *rd/rd* 10-day-old (1 μ g). A single blot was hybridized sequentially with the MBP3, tubulin, and zr.408 probes. The migration of standard RNA markers (BRL) is indicated at the left in kb. METHODS. Total RNA was prepared from C57BL/6J mouse retinas as described²¹, except that caesium trifluoroacetate (Pharmacia) was used instead of CsCl for the density gradient ultracentrifugation according to the manufacturer's instructions. Poly(A)⁺ RNA was selected by oligo(dT) chromatography and electrophoresed in a 1.2% agarose gel containing formaldehyde, transferred to Hybond N+ (Amersham), hybridized at 65 °C as described in Fig. 1, and washed at 0.3 $\times$ SSC/0.1% SDS at 57 °C. The zr.408 probe is described in Fig. 1; MBP3 (Fig. 1a) was amplified for use from pBluescript SK(-)/MBP3 with primers 5'-TTCTACTGTACAGGCCTAT and 5'-GGATCAGCTGCTCGTGCT, corresponding to bases 780–800 and 1,971–1,954 in Fig. 1b, respectively, using standard PCR conditions for *Taq* polymerase (Perkin Elmer-Cetus). A 1.2-kb *Eco*RI fragment of α -tubulin (courtesy of W. Salser²²) was used as a hybridization control for RNA loading.

distribution is distinctly higher in *rd* retinas, suggesting a mutation lying in the PDE β gene. The cause of the increased size of the *rd* mRNA is not yet clear. There are no major insertions or deletions in the *rd* gene compared with the wild-type gene; we are now seeking the site of mutation in the roughly-50-kb region encompassing the gene.

The near identity of the sequence and detection of common targets on genomic DNA and retinal RNA blots establishes that zr.408 is a retinal PDE β cDNA. In conjunction with mapping of zr.408 to the *rd* locale on chromosome 5 (ref. 16), these data support the conclusion that the normal counterpart of the *rd* gene encodes the cGMP-PDE β subunit. We acknowledge the possibility that the mouse PDE β gene is in fact tightly linked, but not identical, to the *rd* locus, although it would be difficult to account for the biochemical data and abnormal PDE β transcript in the *rd* mouse retina. Confirmation awaits results from our ongoing transgenic studies in which we target a normal copy of the gene to the photoreceptor cells of *rd* mice to achieve correction of the disease phenotype. □

Received 13 July; accepted 22 August 1990.

1. Sidman, R. L. & Green, M. C. *J. Hered.* **56**, 23–29 (1965).
2. Dräger, U. C. & Hubel, D. H. *J. comp. Neurol.* **180**, 85–114 (1978).
3. La Vail, M. M. & Sidman, R. L. *Arch. Ophthalmol.* **91**, 394–400 (1974).
4. Farber, D. B. & Lolley, R. N. *Science* **186**, 449–451 (1974).
5. Farber, D. B. & Lolley, R. N. *J. Cyclic Nucleotide Res.* **2**, 139–148 (1976).
6. Bowes, C., Danciger, M., Kozak, C. A. & Farber, D. B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9722–9726 (1989).
7. Lipkin, V. M. et al. *J. biol. Chem.* **265**, 12955–12959 (1990).
8. Yau, K. W. & Baylor, D. A. *A. Rev. Neurosci.* **12**, 289–327 (1989).
9. Bowes, C. & Farber, D. B. *Expl Eye Res.* **45**, 467–480 (1987).
10. Bowes, C., van Veen, T. & Farber, D. B. *Expl Eye Res.* **47**, 369–390 (1988).
11. Farber, D. B., Park, S. & Yamashita, C. *Expl Eye Res.* **46**, 3363–3374 (1988).
12. Baehr, W., Devlin, M. J. & Applebury, M. L. *J. biol. Chem.* **254**, 11669–11677 (1979).
13. Deterre, P., Bigay, J., Robert, M. & Chabre, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2424–2428 (1988).
14. Danciger, M., Kozak, C. A., Li, T., Applebury, M. L. & Farber, D. B. *Expl Eye Res.* **35**, 185–189 (1990).
15. Danciger, M., Tuteja, N., Kozak, C. A. & Farber, D. B. *Expl Eye Res.* **48**, 303–308 (1989).
16. Danciger, M., Bowes, C., Kozak, C. A., LaVail, M. M. & Farber, D. B. *Invest. Ophthalmol. Vis. Sci.* **31**, 1427–1432 (1990).
17. Li, T., Volpp, K. & Applebury, M. L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 293–297 (1990).
18. Maniatis, T., Fritsch, E. G. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1982).

19. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
20. Devereux, J., Haerberli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).
21. Chirgwin, J. A., Przbyla, A., McDonald, R. & Rutter, W. J. *Biochemistry* **18**, 5924–5299 (1979).
22. Davis, R. C. et al. *Devl Biol.* **119**, 164–174 (1987).

ACKNOWLEDGEMENTS. We thank A. L. Rapoport and M.-I. Benito for technical assistance, and J. G. Flannery, K. Volpp and T. Kirkpatrick for help in assembling the data and manuscript. We thank R. Sidman and M. Tang of Harvard Medical School for supplying *rd* mouse strains used in M.L.A. laboratory. The breeding and distribution of these mutants is supported by the National Eye Institute. The work was supported by the George Gund/Retinitis Pigmentosa Foundation and the National Eye Institute. M.L.A. is the recipient of a Jules and Doris Stein Research to Prevent Blindness Professorship. The sequence reported in this paper has been deposited in the EMBL/GenBank GenBank data base, accession no. X55968.

Distinct nuclear and spindle pole body populations of cyclin-cdc2 in fission yeast

Caroline E. Alfa*, Bernard Ducommun†‡, David Beach† & Jeremy S. Hyams*§

* Department of Biology, University College London, Gower Street, London WC1E 6BT, UK

† Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

‡ Present address: EMBL, Heidelberg 6900, FRG

§ To whom correspondence should be addressed

CYCLINS, as subunits of the protein kinase encoded by the *cdc2* gene are major controlling elements of the eukaryotic cell cycle^{1–3}. The fission yeast *Schizosaccharomyces pombe* has a B-type cyclin, which is a nuclear protein^{7,8} encoded by the *cdc13* gene^{4–6}. Here we demonstrate the presence of two spatially distinct *cdc13* cyclin populations in the nucleus of *S. pombe*, one of which is associated with the mitotic spindle poles. Both populations colocalize with the product of the *cdc2* gene (p34^{cdc2}). Treatment of cells with the antimicrotubule drug thiabendazole prevents cyclin degradation and blocks the tyrosine dephosphorylation and activation of *cdc2*. These results suggest a key regulatory role of the *cdc2*-cyclin complex in the initiation of mitotic spindle formation and also that mitotic microtubule function is required for *cdc2* activation.

We have previously shown that *cdc13* in *S. pombe* increases in abundance during interphase (Fig. 1a, b) and disappears rapidly at mitosis^{7,8}. A more detailed investigation of the early stages of mitosis revealed the presence of two bright dots of *cdc13* fluorescence at the nuclear periphery (Fig. 1c). Where the dots were apposed, a second area of diffuse fluorescence within the nucleus was also apparent (Fig. 1c, d, e). When, however, the paired dots were located on opposite sides of the

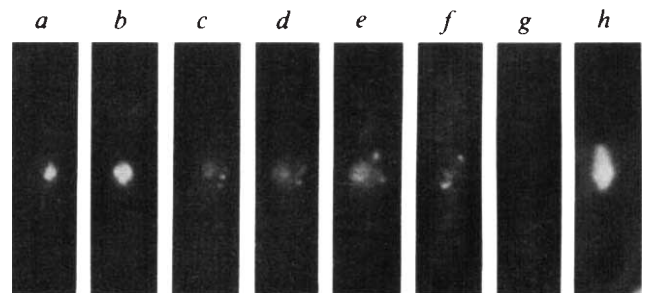
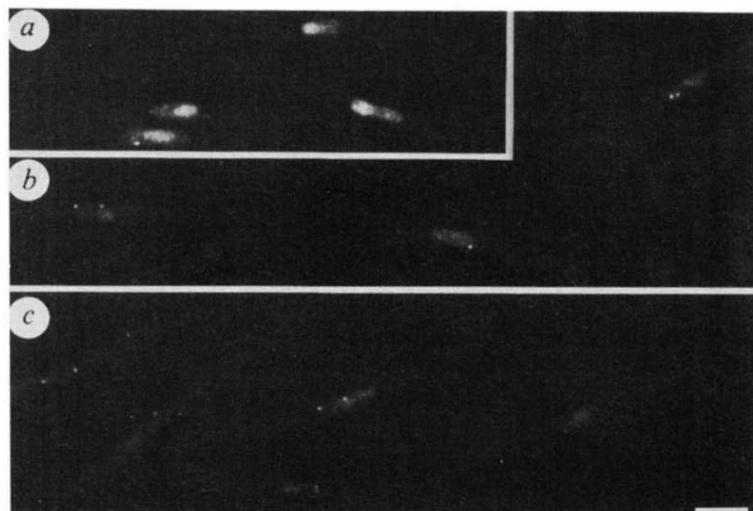


FIG. 1 *cdc13* cyclin indirect immunofluorescence microscopy of wild-type *S. pombe*. Cells from an asynchronous population were prepared using methanol fixation⁷ and stained with an antiserum to *cdc13*, diluted 1/50. a–g, Individual cells stained with anti-*cdc13*; h, 4',6'-diamidino-2-phenylindole (DAPI)/phase contrast image of the cell in g, the onset of anaphase is indicated by the ellipsoid shape of the nucleus. Samples were viewed using a Zeiss Axiophot photomicroscope. Magnification, 2,500 $\times$.

FIG. 2 Staining of the mutant *cdc25.22* with antisera to *cdc13* or *cdc2*. Cells were grown to early exponential phase at 25°C and arrested at the restrictive temperature of 36 °C for 4.25 h (two generations). They were then returned to the permissive temperature (25 °C) and allowed to progress into mitosis⁷. Samples were removed and processed for immunofluorescence microscopy⁷ as follows: *a*, *cdc13* staining at the point of release from temperature arrest; *b*, *cdc13* staining 40 min after release; *c*, *cdc2* staining 20 min after release. Scale bar, 5 µm.



nucleus, this second component was barely detectable or absent (Fig. 1*f*). In contrast, the punctate fluorescence persisted until the onset of chromosome separation (Fig. 1*g, h*). Identical staining patterns were obtained using an antibody against *cdc2* (not shown).

The brevity of the early events of mitosis in wild-type *S. pombe* prompted us to take advantage of the high division synchrony in the mutant *cdc25.22* following temperature arrest and release^{7,9,10}. Arrested cells contained strong *cdc13* fluorescence in the nonchromatin domain of the nucleus (Fig. 2*a*; ref. 7). At mitosis the overall level of fluorescence dramatically decreased and one or two discrete dots per nucleus together with a separate area of fainter fluorescence were again observed (Fig. 2*b*). A similar result was obtained with an anti-*cdc2* antiserum (Fig. 2*c*). Double staining using a combination of *cdc13*- and tubulin-specific antibodies showed that the dots corresponded exactly to the positions of the mitotic spindle poles (Fig. 3*a, b*). The population not associated with the spindle poles diminished as spindle length increased from 1 to 4 µm, with pole staining remaining prominent. Cells containing longer mitotic spindles with associated astral microtubules (Fig. 3*c*) possessed no polar *cdc13* fluorescence (Fig. 3*d*).

The association of both *cdc13* and *cdc2* with the spindle poles suggests a role for these proteins in the initiation of mitotic spindle formation. This was further investigated using the microtubule inhibitor thiabendazole¹¹⁻¹³. Arrested *cdc25.22* cells released into thiabendazole failed to enter mitosis (Fig. 4*a*).

This effect was fully reversible on removal of the drug (not shown). *cdc13* and *cdc2* immunofluorescence remained high (Fig. 4*b*). Further, the rise in *cdc2* histone H1 kinase activity (Fig. 4*c*) and the tyrosine dephosphorylation of *cdc2* (Fig. 4*d*), normally associated with mitotic activation on release of *cdc25* (refs 7, 14), failed to occur.

Our results suggest that *S. pombe* contains two subpopulations of the *cdc2*-*cdc13* kinase which may have distinct roles in the initiation of mitosis. It is unlikely that our antisera recognize a second cyclin because disruption of the *cdc13*⁺ gene results in the abolition of all anti-*cdc13* staining^{7,8}. The fact that the two populations disappear at different times suggests either that a single cyclin can show biphasic degradation at mitosis or that there is a redistribution from the nucleoplasm to the poles before destruction. This first demonstration of a cyclin at the mitotic spindle poles is consistent with reports of *cdc2* at the mammalian centrosome^{15,16}, and further emphasizes the relationship between cyclin and microtubules first inferred from the hypersensitivity of the mutant *cdc13.117* to thiabendazole⁴. Indeed, blocking spindle formation with thiabendazole prevents the degradation and/or redistribution of both *cdc13* and *cdc2* and inhibits activation of the *cdc2* kinase. This strongly suggests that these events are mediated through microtubules or their tubulin subunits. This interpretation becomes even more compelling when coupled to our localization of both *cdc2* and *cdc13* to the spindle pole bodies, the principal microtubule-organizing centres in this yeast^{17,18}. Our experiments do not rule out

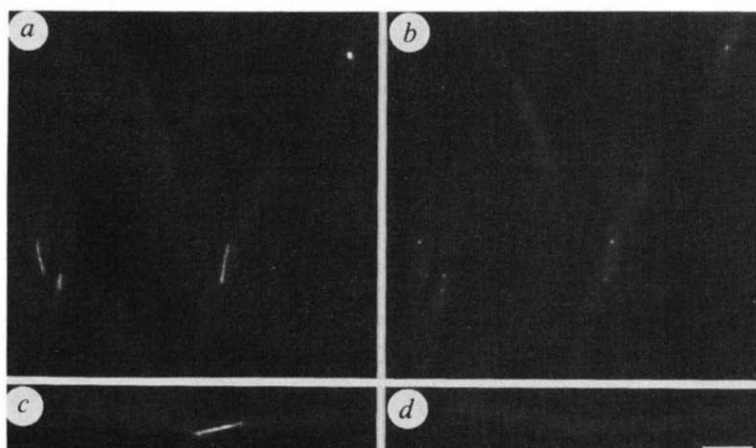


FIG. 3 Double-staining immunofluorescence microscopy of *cdc25.22*. Cells were arrested as described in the legend to Fig. 2 and processed for immunofluorescence⁷ using both anti-*cdc13* and the anti- α -tubulin monoclonal antibody YOL1/34, 20 min after their return to the permissive temperature. Tubulin (*a*) and *cdc13* (*b*) immunofluorescence of a group of cells containing mitotic spindles; tubulin (*c*) and *cdc13* (*d*) staining of a single cell containing a longer spindle with associated astral microtubules. Double staining was achieved by sequential overnight incubations of the cells in antisera in the following order: anti-*cdc13* (1/50 dilution); rhodamine-conjugated anti-rabbit antiserum (ICN Biomedicals); YOL1/34 (1/100 dilution); fluorescein conjugated anti-rat antiserum (ICN Biomedicals). Scale bar, 5 µm.

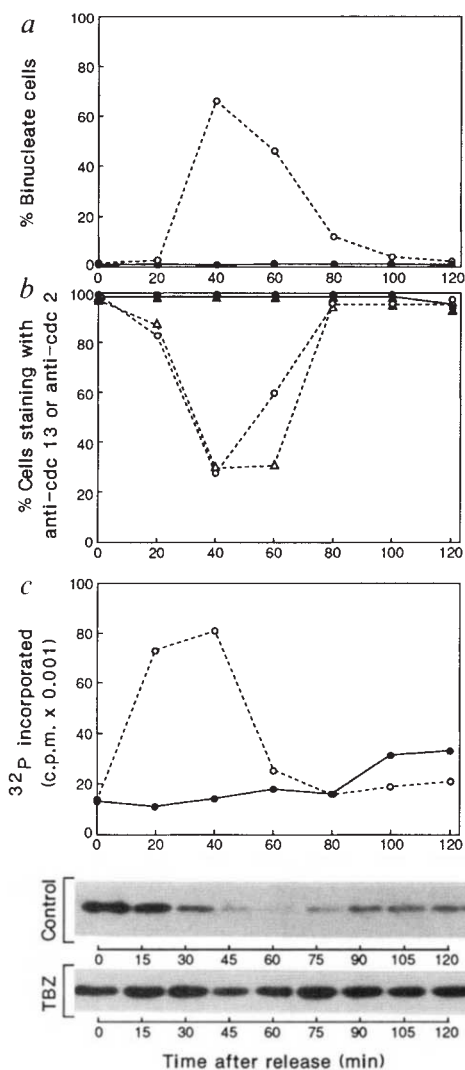


FIG. 4 The effect of thiabendazole on *cdc25.22*, following release from temperature arrest. *a*, A 1-litre culture of cells was arrested as described in the legend to Fig. 2 and shifted back to the permissive temperature. One hour before release of the block, the culture was split in two and thiabendazole (TBZ) was added to a final concentration of $1.00 \mu\text{g ml}^{-1}$ to one culture (solid lines) and the same volume of dimethylsulphoxide was added as a control to the other (dotted lines). At 20-min intervals after release, samples were removed from both cultures and the following parameters determined: *a*, mitotic index from DAPI staining; *b*, percentage of cells stained for *cdc13* (●, ○) or *cdc2* (▲, △) by immunofluorescence; *c*, kinase activity of *cdc2*, measured as ^{32}P incorporation into histone H1 using p13^{suc1} precipitates as described⁷, with modifications (B.D. and D.B., manuscript submitted); *d*, anti-phosphotyrosine (ICN Biomedicals) immunoblots of *cdc2* precipitated using p13^{suc1}.

additional non-tubulin targets for thiabendazole, but a related compound, methylbenzimidazole-2-yl-carbamate (MBC) has recently been shown to inhibit activation of *cdc2* kinase in *Physarum*¹⁹. Other microtubule inhibitors have been shown to arrest cells in mitosis with high levels of *cdc2* kinase^{20,21}. These differences suggest a role for tubulin and/or microtubules in both the activation and inactivation of the *cdc2*-cyclin kinase at different stages of mitosis. □

Received 4 June; accepted 24 August 1990.

- Evans, T., Rosenthal, E. T., Youngbloom, J., Distel, D. & Hunt, T. *Cell* **33**, 389–396 (1983).
- Hunt, T. *Curr. Opin. Cell Biol.* **1**, 268–274 (1989).
- Draetta, G. et al. *Cell* **56**, 829–838 (1989).
- Boher, R. & Beach, D. *EMBO J.* **7**, 2321–2327 (1988).

- Solomon, M., Boher, R., Kirschner, M. & Beach, D. *Cell* **54**, 738 (1988).
- Hagan, I., Hayles, J. & Nurse, P. *J. Cell Sci.* **91**, 587–595 (1988).
- Boher, R. N., Alfa, C. E., Hyams, J. S. & Beach, D. H. *Cell* **58**, 485–497 (1989).
- Alfa, C. E., Boher, R., Beach, D. & Hyams, J. S. *J. Cell Sci. Suppl.* **12**, 9–19 (1989).
- King, S. M. & Hyams, J. S. *Can. J. Microbiol.* **28**, 261–264 (1982).
- Hagan, I. M. Thesis, Univ. London (1988).
- Umesono, K., Hiraoka, Y., Toda, T. & Yanagida, M. *Curr. Genet.* **7**, 123–128 (1983).
- Davidse, L. C. A. *Rev. Phytopathol.* **24**, 43–65 (1986).
- Walker, G. M. *J. gen. Microbiol.* **128**, 61–71 (1982).
- Gould, K. L. & Nurse, P. *Nature* **342**, 39–45 (1989).
- Riabowol, K., Draetta, G., Brizuela, L., Vandre, D. & Beach, D. *Cell* **57**, 393–401 (1989).
- Bailly, E., Dorée, M., Nurse, P. & Bornens, M. *EMBO J.* **8**, 3985–3995 (1989).
- Hyams, J. S. & Borisy, G. G. *J. Cell Biol.* **78**, 401–414 (1978).
- Hagan, I. M. & Hyams, J. S. *J. Cell Sci.* **89**, 343–357 (1988).
- Ducommun, B., Tollen, Y., Garès, M., Beach, D. & Wright, M. *J. Cell Sci.* **96**, 683–689 (1990).
- Draetta, G. & Beach, D. *Cell* **54**, 17–26 (1988).
- Moreno, S., Hayles, J. & Nurse, P. *Cell* **58**, 361–372 (1989).

ACKNOWLEDGEMENTS. We thank C. J. McInerney for assistance with the thiabendazole experiments and J. V. Kilmartin for YOL1/34. The work was supported by the Science and Engineering Research Council (U.S.H.) and the US National Institutes of Health (D.B.). D.B. is an investigator of the Howard Hughes Medical Institute.

Binding of yeast $\alpha 1$ and $\alpha 2$ as a heterodimer to the operator DNA of a haploid-specific gene

A. M. Dranginis

Laboratory of Cellular and Developmental Biology,
National Institute of Diabetes and Digestive and Kidney Diseases,
National Institutes of Health, Building 6, Room B1-10, Bethesda,
Maryland 20892, USA

THE mating-type locus (*MAT*) encodes several DNA-binding proteins, which determine the three cell types of *Saccharomyces cerevisiae*: the α and α haploid cell types, and the α/α diploid cell type¹. One of the products of *MAT*, $\alpha 2$, functions in two cell types. In α cells, $\alpha 2$ represses the α -specific genes by binding to the operator as a dimer². In α/α diploid cells, $\alpha 2$ acts with $\alpha 1$, a product of the other *MAT* allele, to repress a different set of genes, the haploid-specific genes^{3,4}. Until now, the nature of the interaction between $\alpha 1$ and $\alpha 2$ was not known, although it had been suggested that $\alpha 2$ may form a heterodimer with $\alpha 1$ (ref. 5). I show, by using proteins synthesized *in vitro*, that $\alpha 1$ and $\alpha 2$ bind the operator of a haploid-specific gene as a heterodimer. The ability of $\alpha 2$ to form both homodimers and heterodimers with $\alpha 1$, each with a different DNA-binding specificity, explains the dual regulatory functions of $\alpha 2$. This is the first example of regulation by heterodimerization among homeobox-containing proteins, a class that includes proteins responsible for the specification of segment identity in *Drosophila*, mammals and other eukaryotes.

To investigate the basis of this combinatorial regulation, I attached the SP6 promoter to the DNA coding sequences for $\alpha 1$ and $\alpha 2$, and used SP6 polymerase to produce RNA transcripts. The RNA was translated in a wheatgerm system using [^{35}S]methionine, and the products of translation were assayed for DNA-binding activity⁶.

The SP6 promoter was attached using the polymerase chain reaction (PCR)⁷ (Fig. 1a). In the case of $\alpha 1$, one of the yeast genes known to contain introns⁸, complementary DNA was used as the PCR template. The predominant species of $\alpha 1$ messenger RNA in yeast has both introns spliced^{8,9}. Genomic DNA was used as a template to synthesize $\alpha 2$, which has no introns.

A denaturing gel displaying the products of translation *in vitro* is shown in Fig. 1b. Each RNA directed the synthesis of a species of ^{35}S -labelled protein consistent with the predicted size of the product. The $\alpha 1$ migrated as a protein of relative molecular mass 16,000 (M_r , 16K). The size of $\alpha 1$, as predicted from its sequence with both introns removed, is 16.8K. The $\alpha 2$ synthesized *in vitro* was 28K in size. The size of native $\alpha 2$ monomer is 24K¹⁰. The $\alpha 1$ and $\alpha 2$ proteins each contain four methionines^{8,11,12}.



FIG. 1 *a*, Oligonucleotide primers used to synthesize SP6 DNA constructs by PCR⁷. Sequences present in **a1** and **α2** mRNA are in bold^{11,12}. The PCR template for **a1** synthesis was cDNA prepared from poly(A)⁺ RNA from the diploid strain M1-800D × Y1Y319 (ref. 25). Control reactions using cDNA prepared from an isogenic strain with a deletion of *MATa1* (ref. 25) or cDNA prepared without reverse transcriptase, produced no PCR product. This indicates that the template for the **a1** PCR reaction was the mRNA product of *MATa1*. Genomic DNA was used as template for other PCR reactions. Each reaction produced a single DNA product of the expected size, which was purified by gel electrophoresis. RNA was synthesized using SP6 polymerase as described²⁶, using 250 ng PCR product in a 25 μl reaction containing G5'-ppp-5'G at 500 μM⁶. *b*, Autoradiograph of a 17.5% polyacrylamide SDS gel showing the products of *in vitro* translation. The amount of RNA per 10 μl translation is indicated above each lane. A wheatgerm translation system with [³⁵S]methionine was used (Promega). After electrophoresis the gel was treated as described in Fig. 2*b*. Lane M shows *M_r* markers (*M_r* × 10⁻³). A typical 10 μl translation of 50 ng **a1** RNA produced 1.8 ng protein, which corresponds to 11 nM in the translation. The **α2** RNA consistently directed the synthesis of less protein than corresponding amounts of **a1** RNA. Therefore, three times as much **α2** as **a1** RNA was used in the co-translations. *c*, Electrophoretic mobility shift assay of DNA binding by ³⁵S-labelled **a1** and **α2** translated *in vitro*. Number 29 indicates the 29-base-pair haploid-specific operator from the *MATα* intergenic region, nucleotides 1,643–1,672 (ref. 13, Fig. 3*b*). The DNA is not labelled. The zero indicates no DNA. Binding reactions are described in Fig. 2*b*. For the lanes

marked **a1** + **α2**, **a1** and **α2** were separately translated and mixed just before the binding reaction. The two proteins were translated in the same reaction for lanes marked **a1**/α2. Predicted isoelectric points of **a1** and **α2** are 10.5 and 10.2, respectively. They would thus not be expected to enter the gel (5% polyacrylamide in 1 × TBE (ref. 27), pH 8.2) unless they were bound to DNA.

Electrophoretic mobility shift assays were used to detect DNA binding. The DNA chosen for the binding studies was a 29-base-pair fragment (Fig. 3*b*) from *MATα*, which is responsible for haploid-specific expression of **a1**, and which confers haploid-specific regulation on a reporter gene¹³. ³⁵S-labelled protein was incubated with unlabelled DNA, then electrophoresed in a nondenaturing gel. Under these conditions, neither **a1** nor **α2** binds DNA alone (Fig. 1*c*). Nor was binding detected when separate translations of **a1** and **α2** were mixed. When the two proteins were co-translated, however, binding was detected by the presence of a ³⁵S-labelled band in the gel. When the experiment was performed using ³²P-labelled DNA and unlabelled proteins in the presence of co-translated **a1** and **α2**, the ³²P-labelled band moved up to a single band coinciding exactly with the ³⁵S-labelled band in the same gel (data not shown). This showed that the unique band represents a DNA-protein complex.

To determine the stoichiometry of binding, I made an **α2** deletion. Amino acids 2–20 at the amino terminus (Fig. 2*a*) were deleted using PCR with primers 5 and 6 (Fig. 1). This deletion retains *in vitro* binding activity, but lacks the ability to repress *in vivo*¹⁴; it *trans*-inactivates both wild-type **α2** and **a1**/α2 *in vivo*, indicating that the sequences necessary for interaction with **a1** are still intact¹⁴.

When equal amounts of full-length **α2** RNA and RNA of the

shortened form of **α2** were cotranslated with **a1** and bound to the operator DNA, two bands of equal intensity were produced (Fig. 2*b*). The upper band has the same mobility as the band produced by full-length **α2**; the lower band corresponds to the complex with the shortened form of **α2**. If two molecules of **α2** protein bound to each DNA fragment, one would expect to see a third band, intermediate in mobility, representing a complex with one full-length and one shortened form of **α2**^{15,16}. No intermediate band was seen whether **a1** was present in limiting amounts (Fig. 2*b*, lane 4) or in excess (lane 5). This result indicates that only one **α2** molecule is present in each DNA-protein complex.

Similar experiments were performed with amino-terminal deletions of **a1**. Surprisingly, deletion of only amino acids 2–4 of the 126-amino-acid protein virtually eliminated binding with **α2** to DNA in this system. The carboxy terminus contains the homeobox, which is probably the DNA-binding domain¹⁷. As determination of the number of **a1** molecules in the complex was therefore not possible by this method, two-dimensional gel electrophoresis of the ³⁵S-labelled proteins was used. In the first dimension, co-translated **a1** and **α2** were incubated with the operator DNA and subjected to nondenaturing gel electrophoresis. The second dimension was an SDS gel, run perpendicular to the first dimension. Figure 2*c* shows that the protein-DNA complex is resolved into two ³⁵S-labelled proteins spots

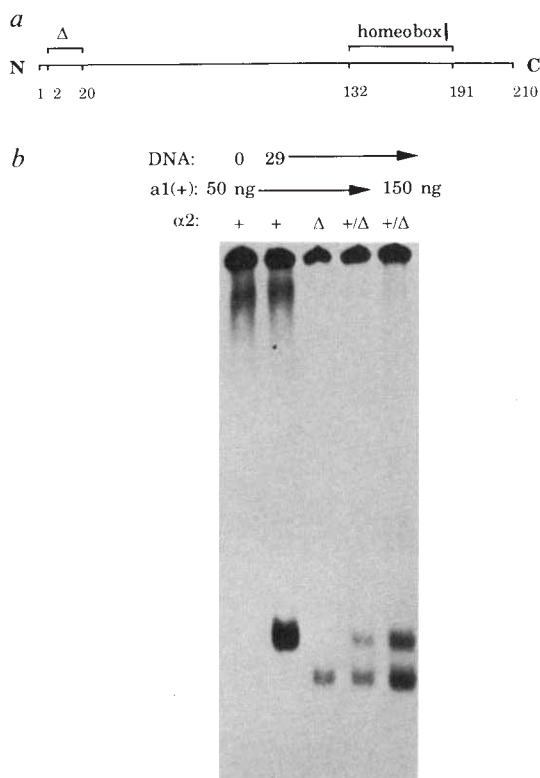
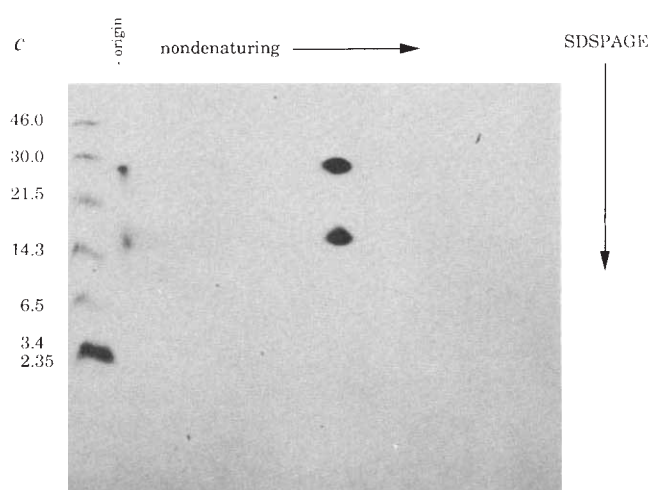


FIG. 2 *a*, Diagram of $\alpha 2$ protein, showing the extent and position of the deletion in $\alpha 2$ $\Delta 2$ -20 relative to the homeobox DNA-binding domain. Numbers mark positions of amino acids. *b*, Electrophoretic mobility shift assay of DNA-binding by the $\alpha 2$ deletion. Wild-type **a1** RNA was cotranslated with either full-length $\alpha 2$ (+), or the $\alpha 2$ $\Delta 2$ -20 deletion (Δ), or an equal mixture



of full-length and deleted $\alpha 2$ RNA (+/ Δ). Each translation (described in Fig. 1b) contained a total of 150 ng $\alpha 2$ RNA. The indicated translation (5 μ l) was added to 2 μ g poly(dI) · poly(dC) in 15 μ l binding buffer, which contained 50 mM NaCl, 10 mM Tris buffer, pH 7.4, 1 mM $MgCl_2$, 1 mM EDTA, 1 mM dithiothreitol, 5% glycerol. Two ng DNA (described in Fig. 1c) was added, and incubated for 20 min at room temperature. Gel loading buffer (5 μ l) contained 1.5 $\times$ binding buffer and 1 mg ml⁻¹ bromophenol blue. The gel (described in Fig. 1c) was fixed with methanol/acetic acid, soaked in autoradiography enhancer, dried and exposed to X-ray film. *c*, Two-dimensional gel electrophoresis of the **a1**/ $\alpha 2$ -DNA complex. Cotranslated ³⁵S-labelled **a1** and $\alpha 2$ were bound to operator DNA as described in *b*, and electrophoresed in a nondenaturing 5% polyacrylamide tube gel in the direction shown. The tube gel was incubated in two changes of SDS sample buffer and affixed to a 17.5% polyacrylamide SDS gel with 5% stacking gel. Electrophoresis in the second dimension was at right angles to the first. The gel was treated as described in *b*. M_r markers ($M_r \times 10^{-3}$) are shown on the left.

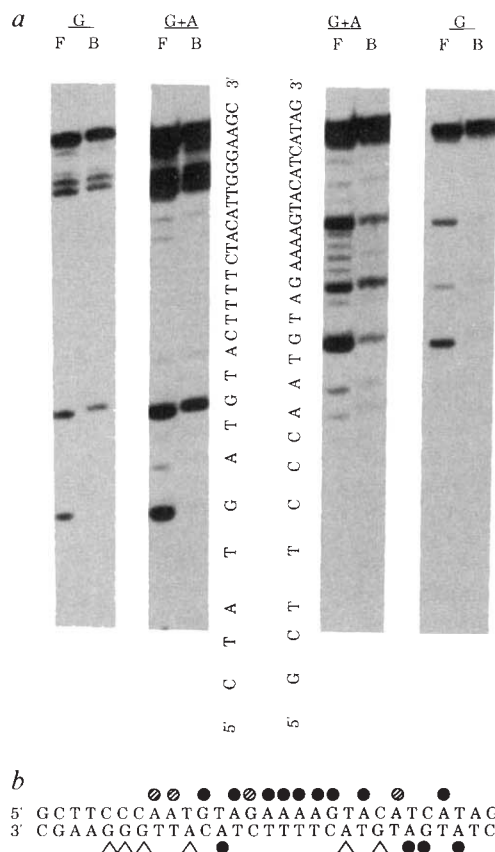
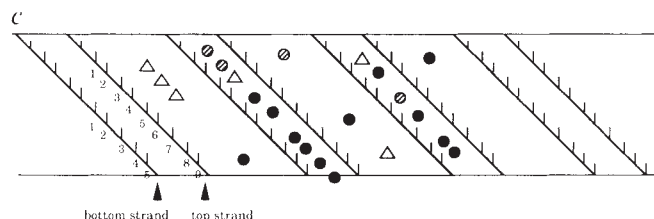


FIG. 3 Methylation interference with DNA binding to cotranslated **a1**/ $\alpha 2$. The experiment was performed as described²⁸. Operator DNA described in Fig. 1c was ³²P-labelled at one end using kinase, and methylated with dimethyl sulphate. The DNA was bound to cotranslated unlabelled **a1**/ $\alpha 2$, and electrophoresed in a nondenaturing gel. Bands containing free (F) and bound (B) DNA were located and eluted from the gel. The DNA was cleaved at methylated positions with piperidine, separated on a 20% sequencing gel, and exposed to X-ray film. Free DNA should be methylated equally at all positions. On the other hand, DNA that has been methylated at positions important for protein binding should be excluded from the bound DNA. *a*, Each of the two strands of the operator is aligned with its sequencing gel pattern. The outer lanes (G) are shorter exposures of the gel, during which the G residues (which are the primary targets of methylation) are revealed. Longer exposures allowed the A residues to appear also (G+A). Bands missing from the bound DNA (lanes B) identify residues where methylation interferes with binding. *b*, Data summarized on the 29-base-pair operator. ●, Bases important for binding to **a1**/ $\alpha 2$. Hatched circles show bases at which methylation partially interferes with binding. △, Bases where methylation does not interfere with binding. *c*, Split helical projection of bases important for binding to **a1** and $\alpha 2$. Symbols, as in *b*, have been positioned on a double helix which is displayed as though it were split along one face and flattened. Numbering of the base pairs is from left to right along the operator sequence shown in *b*. G symbols are shown in the major groove and the A symbols are shown in the minor groove, which are the known sites of modification of these bases by dimethyl sulphate. The pattern of contacts made by **a1** and $\alpha 2$ is made up of two lobes connected at the G which is residue 19. The lobes wrap most of the way around the helix, leaving an irregular channel on one side in which methylation does not interfere with binding.



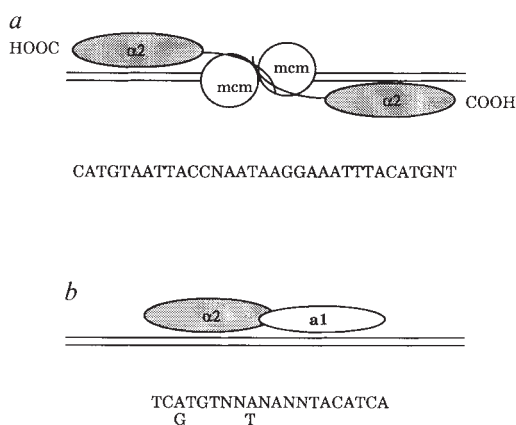


FIG. 4 Two modes by which $\alpha 2$ repressor regulates two different sets of genes. *a*, $\alpha 2$ bound to the operator of an *a*-specific gene. A dimer of $\alpha 2$ binds two half-sites situated 25 base pairs apart, on opposite faces of the helix¹⁰. A dimer of MCM1 occupies the centre of the operator^{18–20}. A consensus $\alpha 2$ /MCM1 sequence is shown, which was derived by Jarvis *et al.* from the *a*-specific genes *STE2*, *STE6*, *BAR1*, *Mfa1* and *MFAa2* (ref. 29). *b*, $\alpha 2$ bound with *a1* to the operator of a haploid-specific gene. The two proteins are shown binding as a heterodimer predominantly to a single face of the helix, although they wrap part of the way around the helix. A consensus sequence of *a1*/ $\alpha 2$ operator DNA is shown, which was derived by Miller *et al.* from the haploid-specific genes *MAT α 1*, *STE5* (two repeated *a1*/ $\alpha 2$ elements) and *HO* (10 elements)⁵. The centre of this sequence, unlike that of the $\alpha 2$ /MCM1 operator, is not conserved. This supports the arguments that *a1* and $\alpha 2$ are sufficient to bind the haploid-specific operator, and that *a1* does not simply substitute for MCM1. The protein contacts seem to wrap part of the way around the helix (shown in Fig. 3c), perhaps in a way similar to that of λ repressor, which uses a flexible region of the protein to wrap around the DNA (for review see ref. 30). Alternatively, the DNA may wrap around the protein complex. The *a1*/ $\alpha 2$ operator used here, like most haploid-specific operators⁵, contains a short run of (dA) · (dT). Such short runs of A or T longer than three base pairs have been proposed to facilitate the wrapping of DNA around protein (for review see ref. 31).

in the second dimension. The upper spot has the mobility of $\alpha 2$; the lower spot, of *a1*. This demonstrates that *a1* is actually present in the DNA–protein complex. Densitometric quantitation of the spots in duplicate gels revealed a ratio of *a1* to $\alpha 2$ of 0.8 ± 0.1 (s.e.m.). One can conclude that each DNA–protein complex contains one molecule of $\alpha 2$ and one of *a1*.

Methylation interference was used to investigate which DNA bases make contact with the *a1*/ $\alpha 2$ complex. These data are summarized in Fig. 3c, which shows the DNA–protein contacts positioned on the double helix. This pattern is quite different from the one produced by $\alpha 2$ at its operator at the *a*-specific genes¹⁰. At the *a*-specific genes, two $\alpha 2$ molecules bind cooperatively with a second, non-cell-type-specific protein, MCM1^{18–20}. A dimer of MCM1 occupies the centre of the $\alpha 2$ operator^{18,19}, whereas the $\alpha 2$ dimer binds to the ends of the operator, on opposite faces of the double helix, spanning MCM1¹⁰ (Fig. 4a).

The stoichiometry of binding demonstrated here excludes a model in which *a1* substitutes for MCM1 at the haploid-specific genes. Figure 4b illustrates a model of $\alpha 2$ binding with *a1* which is derived from this work. A heterodimer of *a1* and $\alpha 2$ is shown binding cooperatively, predominantly to a single face of the helix.

The formation of multiple types of heterodimers is important in the regulation by several of the proteins that contain leucine zippers and the helix–loop–helix family of proteins (see ref. 21 for review). The *a1* and $\alpha 2$ proteins are members of the class of DNA-binding proteins that contains a homeobox²². The homeobox-containing proteins have critical roles in development in other fungi²³ as well as in *Drosophila*, frogs and mammals (see ref. 24 for review). The regulatory activities of many

of these proteins are known to be influenced by the presence of other homeobox-containing proteins. Heterodimerization may be one means by which these combinatorial interactions occur. A few DNA-binding proteins could thus generate many different regulatory specificities. □

Received 21 June; accepted 3 August 1990.

1. Herskowitz, I. *Nature* **342**, 749–757 (1989).
2. Johnson, A. D. & Herskowitz, I. *Cell* **42**, 237–247 (1985).
3. Strathern, J., Hicks, J. & Herskowitz, I. *J. molec. Biol.* **147**, 357–372 (1981).
4. Goutte, C. & Johnson, A. D. *Cell* **52**, 875–882 (1988).
5. Miller, A. M., Mackay, V. L. & Nasmyth, K. A. *Nature* **314**, 598–603 (1985).
6. Hope, I. A. & Struhl, K. *Cell* **43**, 177–188 (1985).
7. Saiki, R. K. *et al.* *Science* **239**, 487–491 (1988).
8. Miller, A. M. *EMBO J.* **3**, 1061–1065 (1984).
9. Ner, S. S. & Smith, M. *Molec. cell. Biol.* **9**, 4613–4620 (1989).
10. Sauer, R. T., Smith, D. L. & Johnson, A. D. *Genes Dev.* **2**, 807–816 (1988).
11. Astell, C. R. *et al.* *Cell* **27**, 15–23 (1981).
12. Tatchell, K., Nasmyth, K. A., Hall, B. D., Astell, C. & Smith, M. *Cell* **27**, 25–35 (1981).
13. Siliciano, P. G. & Tatchell, K. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2320–2324 (1986).
14. Hall, M. & Johnson, A. D. *Science* **237**, 1007–1012 (1987).
15. Smith, D. R., Jackson, I. J. & Brown, D. D. *Cell* **37**, 645–652 (1984).
16. Hope, I. A. & Struhl, K. *EMBO J.* **6**, 2781–2784 (1987).
17. Muller, M. *et al.* *EMBO J.* **7**, 4299–4304 (1988).
18. Keleher, C. A., Goutte, C. & Johnson, A. D. *Cell* **53**, 927–936 (1988).
19. Passmore, S., Elble, R. & Tye, B.-K. *Genes Dev.* **3**, 921–935 (1989).
20. Keleher, C. A., Passmore, S. & Johnson, A. D. *Molec. cell. Biol.* **9**, 5228–5230 (1989).
21. Jones, N. *Cell* **61**, 9–11 (1990).
22. Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., DeRobertis, E. M. & Gehring, W. J. *Nature* **310**, 70–71 (1984).
23. Schulz, B. *et al.* *Cell* **60**, 295–306 (1990).
24. Gehring, W. J. & Hiromi, Y. *A. Rev. Genet.* **20**, 147–173 (1986).
25. Dranginis, A. M. *Molec. cell. Biol.* **9**, 3992–3998 (1989).
26. Melton, D. A. *et al.* *Nucleic Acids Res.* **12**, 7035–7056 (1984).
27. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1989).
28. Ausubel, F. M. *et al.* (eds) *Current Protocols in Molecular Biology*, 12.3.1–12.3.6 (Wiley, New York, 1989).
29. Jarvis, E. E., Hagen, D. C. & Sprague, G. F. Jr. *Molec. cell. Biol.* **8**, 309–320 (1988).
30. Pabo, C. O. & Sauer, R. T. *A. Rev. Biochem.* **53**, 293–321 (1984).
31. Travers, A. A. *Rev. Biochem.* **58**, 427–452 (1989).

ACKNOWLEDGEMENTS. I thank R. T. Simpson, in whose laboratory this work was performed, for his advice and support, A. Hinnebusch, R. T. Simpson, S. Y. Roth, R. H. Morse and A. Wolffe for critical reading of the manuscript, and C. Szent-Gyorgyi for oligonucleotide synthesis.

Sequence and domain structure of talin

D. J. G. Rees^{*†}, Sarah E. Ades^{*†}, S. J. Singer[‡] & Richard O. Hynes^{*†}

^{*} Center for Cancer Research and [†] Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[‡] Department of Biology, B-022, University of California San Diego, La Jolla, California 92093, USA

TALIN is a high-molecular-weight cytoskeletal protein concentrated at regions of cell–substratum contact¹ and, in lymphocytes, at cell–cell contacts^{2,3}. Integrin receptors are involved in the attachment of adherent cells to extracellular matrices^{4,5} and of lymphocytes to other cells⁶. In these situations, talin codistributes with concentrations of integrins in the cell surface membrane^{3,7–9}. Furthermore, *in vitro* binding studies suggest that integrins bind to talin, although with low affinity¹⁰. Talin also binds with high affinity to vinculin¹¹, another cytoskeletal protein concentrated at points of cell adhesion¹². Finally, talin is a substrate for the Ca^{2+} -activated protease, calpain II^{13,14}, which is also concentrated at points of cell–substratum contact¹⁴. To learn more about the structure of talin and its involvement in transmembrane connections between extracellular adhesions and the cytoskeleton, we have cloned and sequenced murine talin. We describe a model for the structure of talin based on this sequence and other data. Homologies between talin and other proteins define a novel family of submembranous cytoskeleton-associated proteins all apparently involved in connections to the plasma membrane.

```

1  MVALSLKISIGNVVKTMQFEPSTMVYDACRMIRERIPEALAGPPNDFGLFLSDDDPKKGIWLEAGKALDYMLRNGDTMEYRKQORPLKIRMLDGTVKTI
101 MVDDSKTVDMLMTICARIGITNHDEYSLVRELMEKKDDEGTGLRKDKTLRDEKKMEKLLKQLHTDDELNLWDHGRTLREQGVVEHETLLLRKPFYS
201 DQNVSDRPVQLNLLYVQARDDILNGSHPVSFDKACEFAGFQCQIQFGPHNEQKHKAGFLDLKDFLPKEYVVKQKGERKIFQAHKNCQGMSEIAKVRYVK
301 LARSLKTYGVSFVLVKEKMKGNKLVPRLLGITKECVMRVDEKTEKVIQEWSLTNIKRWAAAPKSFITLDFGDDYQDGYYSVQTTEGEQIAQLIAGYIDITL
401 KKKKSKDHFGLLEGDEESTMLEDSVSPKSTVLQQQYNRVGKVEHGSVALPAIMRSGASGSPENFQVGSMPPAQQQITSGQMRGHMPPLTSAQQALTGTIN
501 SSMQAVQAAQATLDDFETLPPLGQDAASKAWRNKMDSEKHEIHSQVDAITAGTASVVNLTAGDPAETDITAVGCAVTTISSNLTEMSRGVKLLAALLED
601 EGGNGRPLLQAAKGLAGAVSELLRSAQPASAEPRONLLQAAAGNVGQASGELLQQIGESDTPHFDQVLMQLANAVASAAAAVLVLAQSAVQRTEDSGLQT
701 QVIAAATQCALSTSQLVACTKVVPAPTISPVQCEQLVEAGRLVAKAVEGCVSASQAAATEDGQLLRGVGAAATAVTQALNELLQHVKAHATGAGPAGRYDQ
801 ATDTILTVTENIFSSMGDAGEMVRQARILAQATSDLVNAIKADAEGESDLENSRKLLSAAKILADATAMVEAAKGAAPDSEEQQRRLREAEEGLRMA
901 TNAAQNAIKKKLVORLEHAAKQAAASATQTIAAAQHAASAPKASAGPOPLLVSCKAVAEQIPLLVQGVRSQAQPDSPSAQLALIAASQSFLQPGGKM
1001 VAAAKASVPTIQDQASAMQLSQCAKNLGTALAEELTAQAQAQAEACGPLEMSALSVMQNEKDLQEIKAARDGKLLPLPGTMEKCTQDLGNSTKAVSS
1101 AIAKLLGEIAQGNENYAGIAARDVAGGLRSLAQAARGVAAALTSDPAVQAIVLDTASDVLDKASSLIEEAKKASGHPGDPESQORLAQVAKAVTQALNRCV
1201 SCLPGQRVDNALRAVGDASKRLSDLLPSTGTQEAQSRLNEAAAGLNCQAATELVQASRGTPQDLARASGRFGQDFSTFLEAGVEMAGQAPSQEDRAQ
1301 VVSNLKGISMSSSKLLLAAKALSTDPASPNLKSQLAAARAVTDSINQLITMCTQQAPGQKECDNALRQLETVRELLENPVQPINDMSYFGCLDSVMENS
1401 KVLGEAMTGISQNAKNGNLPFEGDAIATASKALCGFTEAAQAAYLVGVSDPNQAGQQGLVEPTQFARANQAIQMACQSLGEPGCTQAQVLSAATIVAK
1501 HTSALCNSCRLASARTANPTAKRFVQSAKEVANSTANLVKTIKALDGDFTENRAQCRAAATAPLLEAVDNLSAFASNPEFSSVPAQISPEGRAAMEPIV
1601 ISAKTMLESAGGLIQTARALAVNPRDPPRWSVLGHSRTVSDSIKKLITSMRDKAPGQLECEATAAALNSCLRDLDQASLAAVVSQQLAPREGISQEALHT
1701 QMLTAVQEISHLIEPLASAAARAEASQLGHKVSQMAQYFEPLTLAAVGAASKTSLSHPPQMMALLDQTKTLAESALQLLYTAKEAGGNPKQAAHTQEALAEAV
1801 QMMTEAVEDLTTLTNEAAASAGVVGGMVDSITQAINQLDEGPMGDPESFVDYQTTMVRTAKAIAVTQEMVTKSNTSPEELGPLANQLTSDYGRLASQA
1901 KPAAVAAENEEIGAHIKHRVQELGHGCSALVTKAGALQCSPSDVYTKKELIECARRVSEKVSHVLAALQAGNRGTQACITAAASAVSGIIADLDTTIMFAT
2001 AGTLNREGAETFADHREGILKTAKVLVEDTKVLVQNAAGSQEKLAQAQSSVATITRLADVVKLGAAASLGAEDPETQVVLINAVKDVAKALGDLISATKA
2101 AAGKVGDDPAVWQLKNSAKVMVTNVTSLKTVKAVEDEATKGTRALEATTEHIRQELAVFCSPEPPAKTSTPEDFIRMKGITMATAKAVAAAGNSCRQED
2201 VIATANLSRRATADMLRACKEAAAFHPEVAPDVRLRALHYGREGANGYLELLDHVLLTLQKPNPDLKQQLTGHSKRKRVAGSVTELIQAAEAMKGTWVDPED
2301 PTVIAENELLGAAAAIEAAAKKLEQLKPRAPKPEADESNFEEQILEAAKSIAAATSALVKAAASAAQRELVAQGVGAIPANALDDGQWSQGLISAAARMV
2401 AAATNNLCEAAANAAVQGHASQEKLISSAKQVAASTAQLLVACKVKADQSEAMKRLQAAAGNAVKRASDNLVKAAQKAAAFEDQENETVVVKEKMGVGGIAQ
2501 ITAAQEEMLRKERELEEARKKLAQIRQQYKFLPSELRDEH 2541

```

The initial talin complementary DNA clone was isolated by antibody screening and validated by western blotting of fusion proteins and of talin and by immunofluorescence studies (data not shown). Antisera raised against the fusion proteins also reacted with talin in similar assays (data not shown, see Fig. 1). Further cDNA clones were then isolated by DNA hybridization to cover the complete sequence and selected clones were sequenced (Fig. 1). The sequence defines an open reading frame encoding 2,541 amino acids (Fig. 1). The position of the N-terminal methionine was confirmed by amino-acid sequencing (see Fig. 1 and below) and is encoded by a nucleotide sequence (CCACCATGG) conforming to the consensus typical of eukaryotic initiation sites¹⁵. A typical polyadenylation signal, AATAAA, precedes a poly(A) tail by 11 nucleotides. Restriction

enzyme analyses and partial sequence data on a large number of overlapping cDNA clones revealed no evidence for variant forms (alternative splicing etc.) other than the polymorphic residues described in Fig. 1.

The sequence is markedly nonhomogenous (Fig. 1) and can be considered in two blocks. The first 600 residues are highly polar (28% charged residues), whereas the last 1,900 residues are highly enriched in alanine (18%) and poor in aromatic residues (2%). The C-terminal 60 residues are highly charged (>30%). The division of the talin sequence into blocks is readily seen in Fig. 2a which shows a comparison of the mouse talin sequence with itself; whereas the first 600 residues show no internal homologies, the last 1,900 show apparent homologies. These homologies are due to the high content of alanine and

◀ FIG. 1 Amino-acid sequence (single-letter notation) of mouse talin deduced from cDNA sequence. The N-terminal sequence was confirmed by amino-acid sequencing of the 43K fragment of chicken gizzard talin which agreed with the deduced mouse sequence at 16 out of 17 positions commencing with residue 2 (underlined). The sequence commencing at residue 434 (underlined) is closely similar to the sequence obtained from the 190K fragment produced by calpain digestion of chicken gizzard talin (identities at 11 out of 18 positions). This sequence defines the point of calpain cleavage (arrow) which is preceded by a segment of protein (italics) that matches so-called PEST sequences, which have been hypothesized to be involved in protease cleavages^{33,34}. The segment homologous with band 4.1 and ezrin (see Figs 2 and 3) is boxed and the alanine-rich nature of the tail segment is highlighted by bolding of runs of alanine residues. The coding region is preceded by 159 nucleotides of 5' untranslated sequence and followed by a 440 nucleotide long 3'-untranslated region (all segments confirmed by analysis of the positional base bias³⁵). The complete nucleotide sequence determined is 8,226 bases long without poly(A) tail, corresponding well with the size of the ~8.5 kilobase (kb) messenger RNA detected on northern blots (data not shown). The total cDNA sequence has been submitted to GENBANK/EMBL/DBJ (accession number: X56123).

METHODS. The initial cDNA clone was isolated from an oligo(dT)-primed λ gt11 cDNA library prepared from poly(A)⁺ RNA isolated from mouse Swiss 3T3 cells, using a polyclonal anti-talin antibody affinity-purified on 210–240K proteins from Swiss 3T3 cells. This clone was confirmed by immunoblotting of the fusion protein, which was then used to affinity-purify antibodies that reacted with chicken talin on immunoblots of total chicken embryo fibroblast protein and stained focal contacts in 3T3 cells and chick embryo fibroblasts in immunofluorescence experiments. Further clones were then isolated from two other libraries, one of cDNA made from Balb/c 3T3 RNA, cloned in λ gt11 (Clontech) and one prepared with cDNA from Balb/c 3T3 RNA using the Librarian system from Invitrogen and cloned in λ ZAPII according to the manufacturer's instructions. Selected clones were subcloned into M13mp18 or M13mp19 as sonicated fragments and sequenced using either the Klenow fragment of DNA polymerase³⁶, or modified T7 DNA polymerase³⁷. Sequence data were assembled and analysed using the programs of Staden³⁵ and UWGCG. Two residues were found to be variant in this analysis; residue 3,473 was either T or C, coding for either leucine or proline at amino-acid 1,105; and residue 6,698 was either A or T, coding for either methionine or lysine at amino acid 2,180. Talin was purified from chicken gizzards¹ and cleaved with calpain II purified from bovine heart³⁸. The 47K and 190K fragments were separated on 5% SDS-polyacrylamide gels and transferred to Immobilon membranes³⁹. Bands were cut out and sequenced in an Applied Biosystems Protein Sequencer.

do not reflect any readily discernible repeating structures. Secondary structure predictions suggest a high content of α helix consistent with the high α -helix content of talin¹⁶, but the sequence is not consistent with a coiled-coil structure. These structural predictions are consistent with electron microscope images of talin¹⁶, which show an elongated molecule (60 nm) with a globular head and a flexible tail.

Also consistent with the suggestion of two structural domains in talin is its cleavage by calpain II into two fragments of apparent relative molecular masses 47,000 and 190,000 (M_r 47K and 190K) (ref. 14). We isolated these fragments from a calpain II digest of chicken gizzard talin and determined their N-terminal amino-acid sequences. Identities with the mouse talin sequence unambiguously identified the 47K fragment as the N terminus of talin and the 190K fragment as the C-terminal domain (Fig. 1). The calpain II cleavage site is before residue 434. The predicted M_r of the 433-residue N-terminal domain is 49,981 consistent with the apparent M_r of 47,000. The predicted M_r of the C-terminal fragment is 219,873, which is larger than its apparent M_r on SDS-PAGE (190K). The same is true for intact talin; actual M_r is 269,854K, apparent M_r is 225–235K. This discrepancy is likely to be a consequence of the unusual amino-acid composition and/or secondary structure of the C-terminal domain of talin.

The mouse talin sequence was compared with the Owl protein sequence database¹⁷ using the FASTA program. The C-terminal tail segment showed only weak relationships with various structural proteins. We attribute these apparent relationships to similarities in secondary structure (such as α helices) rather than to evolutionary relatedness. In particular, there are no significant relationships with other large cytoskeletal proteins with extended coiled α helices (myosin, tropomyosin) or repeating segments of α helix (spectrin, α -actinin, dystrophin). A segment from the N-terminal region of talin showed homology with two other proteins, however, as shown on the DIAGON plots (Fig. 2b and c) and displayed in Fig. 3. These proteins are band 4.1, originally identified in erythrocytes but also present in many other cells^{18,19}, and a protein, known variously as ezrin²⁰, cytovillin^{21,22} or p81^{23,24}, originally described in

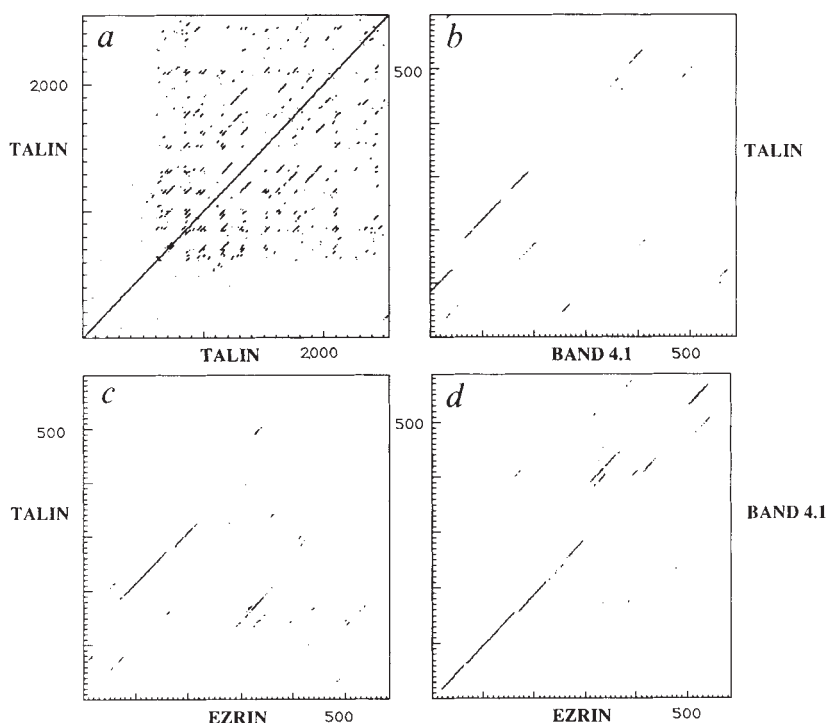


FIG. 2 Comparison of mouse talin amino-acid sequence with itself and with the sequences of human band 4.1 and ezrin/cytovillin/p81. *a*, Self comparison shows that the sequence can be considered in two parts. The first 600 residues are unrelated to the rest of the sequence. The last 1,900 residues show some degree of self-relatedness, but on examination this is largely attributable to the high content of alanine and does not reflect repeated structural units. *b*, *c*, Comparisons with band 4.1 and ezrin show in each case that a segment of talin (roughly residues 150–350) is homologous with the N-terminal 200 residues of each of these proteins. *d*, Comparison of band 4.1 with ezrin shows that these two proteins are homologous over the first 300 residues.

METHODS. Comparisons were made using the DIAGON program of Staden with a window of 51 residues and a threshold of 540.

Human Band 4.1	47	WDNATSKTWLDSAKEIKKQV..RGVPWNFTFNVKFYPDPDA.QLTEDITRYYLCLQLRQDIVAGRLPCSFATLAL	
Mouse Talin	165	LHTDELNWLHDGRTLREQG..VEEHETLLLRKFFYSQNVDSRDPVQNLNLLYVQARDDILNGSHPVSFDKACE	
Human Ezrin	49	VDNKGFPFTWLKLDKKVSAQEVKRNPLQKFRAKFYPEDVAEELIQDITQKLFFLQVKEGILSDEIYCPPEAVL	
X.laevis 4.1	239	WESPTCKVWLDPLKDIRKQV..HGGPCEFTSNVKFYPPDPA.QLSEDI TRYYLCLQLRKDIFSGRLPCSFATLAL	
Human Band 4.1	119	LGSYTIQSELGDYDPELHGVDYVSDFKLAPNQKE.....LEEKVMEHLKHSYRSMTPAQADLEFLENAKKLS	
Mouse Talin	238	FAGFQCQIQFGPHNEQKHKAGFLDLKDFLPKEYVKQKG.....ERKIFQAHKNCGQMSEIEAKVRYVKLARSLE	
Human Ezrin	124	LGSYAVQAKFGDYNKEVHKSGYLSERLIPQVRMDQHKLTRDQWEDRIQVWHAERGLMKDNAMLEYLKIADLE	
X.laevis 4.1	311	LGSYTVQSEVGDYEDLHGVDYVSEFKLSPNQTKD.....LEEKVGEHLKHSYRSMTPAQADLEFLENAKKLT	
Human Band 4.1	186	MYGVDLHKAKDLEG....VDIILGVCSSGLLVYKDKLR.INRFPWPKVLKISYKR..SSFFIKIRPG	245
Mouse Talin	307	TYGVSFFLVKEKMGKNKLVPRLLGITKECVMRVDEKTK.EVIQEWSLTNIKRWAASPKSFTLDFGDY	373
Human Ezrin	199	MYGINYFEIKNKKGTDLWLGVDAALGN...IYEKDDKLTPTKIGFPWSEIRNISFND..KKFVIKPIDK	261
X.laevis 4.1	378	MYGVDIHQAKDLEG....VDIKLGVCSGLMVFKNLR.INRFPWPKVLKISYKR..SSFFIKIRPG	437

FIG. 3 Comparison of the homologous N-terminal domains of talin, band 4.1 and ezrin. Alignment of the sequences of mouse talin, human ezrin²⁴, human band 4.1 (refs 33 and 40) and *Xenopus* band 4.1 (ref. 41). Residues 19–46 in the human erythroid band 4.1 sequence can be alternatively spliced⁴²; the homology between talin and the other proteins follows this alternatively

spliced segment. Identities are marked by lines and conservative substitutions (D, E; K, R; L, I, V, M, C; Y, F, W; S, T) by colons. The region displayed exhibits 20% identity between talin and human band 4.1 (9.6 s.d. above random), 23% identity between talin and ezrin (9.8 s.d. above random) and 35% identity between ezrin and band 4.1 (16.5 s.d. above random).

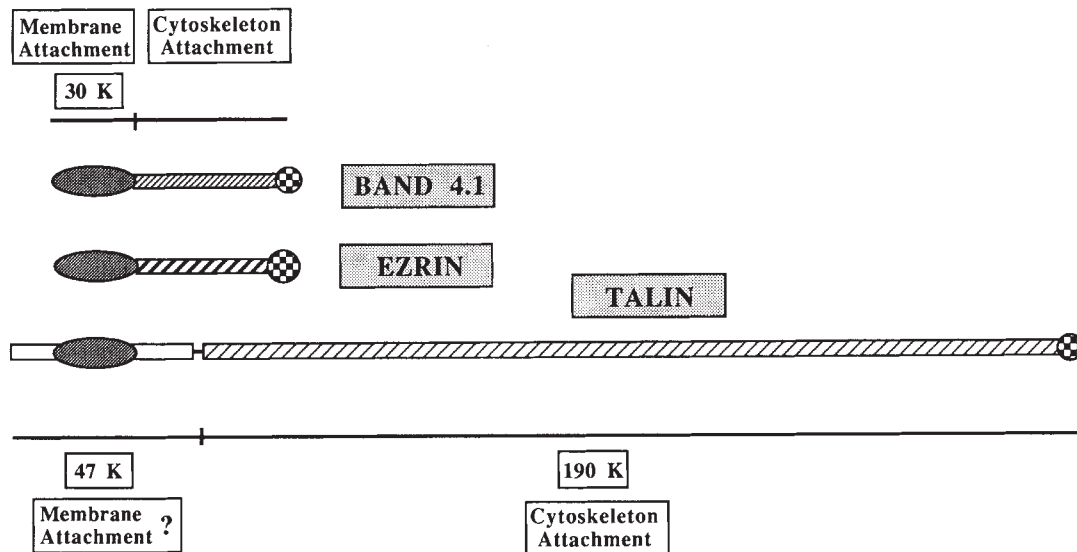
intestinal brush borders²⁵ but also present in many other cell types²⁶. A segment of 200–220 amino acids is homologous in talin, ezrin and band 4.1 (Figs 2 and 3).

The existence of a homologous N-terminal domain in these three proteins seems likely to have a bearing on the interaction of each with protein components of the plasma membrane. Band 4.1 has been analysed most extensively; it links the spectrin–actin cytoskeleton of erythrocytes to the plasma membrane. Band 4.1 is reported to bind two integral membrane proteins, with highest affinity for glycophorin^{27,28} and lower affinity for the anion transporter, band III (ref. 29). The binding site for glycophorin has been mapped to a 30K fragment of band 4.1 that contains the region of homology with talin³⁰. Although the interactions of ezrin with other proteins have not been defined, it is clearly localized in a submembranous region in the brush border of intestinal epithelial cells and more generally in other cells²¹. Ezrin also colocalizes with a variety of cytoskeletal elements^{31,32}.

Thus, both these proteins, like talin, are found at interfaces between the cytoskeleton and the plasma membrane and it seems likely that the domains homologous with the glycophorin-binding domain of 4.1 (see Fig. 3) are involved in membrane attachment. It has been suggested that talin binds to integrins¹⁰. There are no obvious homologies between the cytoplasmic domains of glycophorin and integrin subunits. It is also worth noting that the affinity of band 4.1 for glycophorin is enhanced by phosphorylated derivatives of phosphatidylinositol²⁸; it will be of interest to determine whether the same is true for the binding of talin to integrins. It is not known whether or not this binding occurs through the 47K N-terminal domain, although the homology shown in Fig. 3 might suggest such a model. The availability of cDNA clones encoding this domain should facilitate further analyses of these questions.

Turning next to the C-terminal domains of the three homologous proteins, the available data suggest that they are

FIG. 4 Models of the structures of talin, band 4.1 and ezrin. Each protein has a homologous N-terminal domain (shaded oval) followed by a domain rich in α helix (hatched rectangles; different shading indicates that these domains are not homologous with each other). The C-terminal segments (chequered circles) in each protein are highly charged. Evidence exists for membrane and cytoskeleton binding sites in the indicated domains of band 4.1 and for vinculin binding in the 190K domain of talin. Membrane attachment through the homologous N-terminal domains of each protein is hypothesized.



involved in binding to other cytoskeletal proteins. The spectrin-actin binding site of band 4.1 is located in the C-terminal half of the protein in a segment predicted to be predominantly α -helical³³. The 190K fragment of talin binds to vinculin¹¹ and is also predicted to contain significant α -helical structure. No binding data are available for ezrin; however, it also has a segment predicted to be α -helical. A reasonable argument can thus be made that the three proteins—talin, band 4.1 and ezrin—have analogous structures (Fig. 4). Each molecule has a homologous N-terminal domain, known to be involved in membrane attachment in band 4.1 and hypothesized to be involved in interactions of talin with integrins and of ezrin with unknown membrane protein(s). In each molecule this homologous domain is followed by a segment postulated to be rich in α helix. These α -helical segments are not homologous but, in the case of band 4.1 and talin, there is evidence for functional as well as structural analogy in that each binds to cytoskeletal proteins—spectrin-actin and vinculin, respectively. Finally, each of the three proteins ends in a short highly charged segment.

Thus, the sequence data we present here, together with data from the literature and the homologies we describe, allow us to propose that talin is a member of a family of submembranous cytoskeletal proteins involved in connections of major cytoskeletal structures to the plasma membrane. The homologies and analogies among the proteins in this family suggest many lines for future investigation. □

X-ray structure of phospholipase A₂ complexed with a substrate-derived inhibitor

Marjolein M. G. M. Thunnissen, Eiso AB, Kor H. Kalk, Jan Drenth, Bauke W. Dijkstra*, Oscar P. Kuipers†, Ruud Dijkman†, Gerard H. de Haas† & Hubertus M. Verheij†

Laboratory of Chemical Physics, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands

† Department of Enzymology and Protein Engineering, University of Utrecht, Center for Biomembranes and Lipid Enzymology, PO Box 80054, 3508 TB Utrecht, The Netherlands

* To whom correspondence should be addressed

PHOSPHOLIPASES A₂ play a part in a number of physiologically important cellular processes such as inflammation, blood platelet aggregation and acute hypersensitivity^{1,2}. These processes are all initiated by the release of arachidonic acid from cell membranes which is catalysed by intracellular phospholipases A₂ and followed by conversion of arachidonic acid to prostaglandins, leukotrienes or thromboxanes³. An imbalance in the production of these compounds can lead to chronic inflammatory diseases such as rheumatoid arthritis and asthma. Inhibitors of phospholipase A₂ might therefore act to reduce the effects of inflammation, so structural information about the binding of phospholipase A₂ to its substrates could be helpful in the design of therapeutic drugs. The three-dimensional structure is not known for any intracellular phospholipase A₂, but these enzymes share significant sequence homology^{4–6} with secreted phospholipases, for which some of the structures have been determined^{7–10}. Here we report the structure of a complex between an extracellular phospholipase A₂ and a competitively inhibiting substrate analogue, which reveals considerable detail about the interaction and suggests a mechanism for catalysis by this enzyme.

The inhibitor used in these studies is (R)-2-dodecanoyl-amino-1-hexanol-phosphoglycol (Fig. 1). This inhibitor binds by over three orders of magnitude more strongly than the substrate (R)-1,2-di-dodecanoyl-glycerol-3-phosphocholine¹¹ to phospholipase A₂. A porcine phospholipase A₂ mutant obtained by site-directed mutagenesis was used because of its better crystallization properties. Moreover, this mutant, in common with cellular phospholipases A₂, lacks residues 62–66 and has a high activity on monomeric and aggregated substrates¹². To increase the affinity for monomeric substrates still further, we created a single Trp phospholipase A₂ (W3F,L31W; see Table 1).

The complex of this mutant phospholipase A₂ and the inhibitor was crystallized (space group $P2_12_12_1$, with two

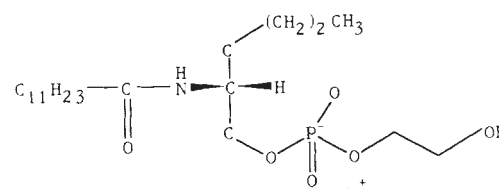


FIG. 1 The inhibitor (R)-2-dodecanoyl-amino-1-hexanol-phosphoglycol used in our investigation. The inhibitor was prepared from *nor*-leucine-methylester as described previously²². Inhibition measured at pH 8 by pH-stat titration using (R)-1,2-di-dodecanoyl-glycerol-3-phosphocholine (3 mM in 3 mM sodium taurodeoxycholate) as substrate¹¹ showed that the inhibitor binds 1,200 times more strongly than a natural substrate molecule.

Received 29 May; accepted 15 August 1990.

- Burridge, K. & Connell, L. *J. Cell Biol.* **97**, 359–367 (1983).
- Kupfer, A., Singer, S. J. & Dennert, G. *J. exp. Med.* **163**, 489–498 (1986).
- Burn, P., Kupfer, A. & Singer, S. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 497–501 (1988).
- Hynes, R. O. *Cell* **48**, 549–554 (1987).
- Ruoslahti, E. & Pierschbacher, M. D. *Science* **238**, 491–497 (1987).
- Anderson, D. C. & Springer, T. A. *Rev. Med.* **38**, 175–194 (1987).
- Chen, W. T., Hasegawa, T., Hasegawa, C., Weinstock, C. & Yamada, K. M. *J. Cell Biol.* **100**, 1103–1114 (1985).
- Damsky, C. M., Knudsen, K. A., Bradley, D., Buck, C. A. & Horwitz, A. J. *J. Cell Biol.* **100**, 1528–1539 (1985).
- Kupfer, A. & Singer, S. J. *J. exp. Med.* **170**, 1697–1713 (1989).
- Horwitz, A., Duggan, E., Buck, C., Beckerle, M. C. & Burridge, K. *Nature* **320**, 531–533 (1986).
- Burridge, K. & Mangeat, P. *Nature* **308**, 744–746 (1984).
- Geiger, B. *Cell* **18**, 193–205 (1979).
- Fox, J. E. B., Goll, D. E., Reynolds, C. C. & Phillips, D. R. *J. Biol. Chem.* **260**, 1060–1066 (1985).
- Beckerle, M. C., Burridge, K., DeMartino, G. N. & Croall, D. E. *Cell* **51**, 569–577 (1987).
- Kozak, M. *Nucleic Acids Res.* **12**, 857–872 (1984).
- Molony, L., McCaslin, D., Abernethy, J., Paschal, B. & Burridge, K. *J. Biol. Chem.* **262**, 7790–7795 (1987).
- Akrigg, D. *et al. Nature* **335**, 745–746 (1988).
- Bennett, V. A. *Rev. Biochem.* **54**, 273–304 (1985).
- Marchesi, V. T. A. *Rev. Cell Biol.* **1**, 531–561 (1985).
- Bretscher, A. *Meth. Enzym.* **134**, 24–37 (1986).
- Pakkanen, R., Hedman, K., Turunen, O., Wahlstrom, T. & Vaheri, A. *J. Histochem. Cytochem.* **35**, 809–816 (1987).
- Turunen, O. *et al. J. Biol. Chem.* **264**, 16727–16732 (1989).
- Gould, K. L., Cooper, J. A., Bretscher, A. & Hunter, T. *J. Cell Biol.* **102**, 660–669 (1986).
- Gould, K. L., Bretscher, A., Esch, F. S. & Hunter, T. *EMBO J.* **8**, 4133–4142 (1989).
- Bretscher, A. *J. Cell Biol.* **97**, 425–432 (1983).
- Pakkanen, R. *J. cell. Biochem.* **38**, 65–75 (1988).
- Anderson, R. A. & Lovrien, R. E. *Nature* **307**, 655–658 (1984).
- Anderson, R. A. & Marchesi, V. T. *Nature* **318**, 295–298 (1985).
- Pasternack, G. R., Leto, T. L., Anderson, R. A. & Marchesi, V. T. *J. Biol. Chem.* **260**, 3676–3683 (1985).
- Leto, T. L., Correia, I., Tobe, T., Anderson, R. A. & Horne, W. C. in *Membrane Skeletons and Cytoskeletal Membrane Associations* (eds Bennett, V., Cohen, C. M., Lux, S. E. & Palek, J.) 201–209 (Liss, New York, 1986).
- Birgbauer, E. & Solomon, F. *J. Cell Biol.* **109**, 1609–1620 (1989).
- Goslin, K., Birgbauer, E., Banker, G. & Solomon, F. *J. Cell Biol.* **109**, 1621–1631 (1989).
- Conboy, J., Kan, Y. W., Shohet, S. B. & Mohandas, N. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9512–9516 (1986).
- Rogers, S., Wells, R. & Rechsteiner, M. *Science* **234**, 364–368 (1986).
- Wang, K. K. W., Villalobo, A. & Roufogalis, B. D. *Biochem. J.* **262**, 693–706 (1989).
- Staden, R. *Meth. Enzym.* **183**, 163–180 (1990).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Tabor, S. & Richardson, C. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4767–4771 (1987).
- DeMartino, G. N. & Croall, D. E. *Biochemistry* **22**, 6287–6291 (1983).
- Matsuda, P. *J. Biol. Chem.* **262**, 10035–10038 (1987).
- Tang, T. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 3713–3717 (1988).
- Giebelhaus, D. H., Eib, D. W. & Moon, R. T. *Cell* **53**, 601–615 (1988).
- Tang, T. K., Qin, Z., Leto, T., Marchesi, V. T. & Benz, E. J. *Jr J. Cell Biol.* **110**, 617–624 (1990).

ACKNOWLEDGEMENTS. We thank M. Adams for the preparation of anti-talin antibodies, D. Croall for advice on the purification of calpain II, R. Cook and the MIT/HMI Biopolymer Laboratory for protein sequencing and T. Hunter for generous support, encouragement and assistance. We also thank E. Pasquale, J. Pines, L. Connell, A. Tomacka and P. Johnson for advice. This work was supported by the US Public Health Service (R.O.H. and S.J.S.) and the Howard Hughes Medical Institute (R.O.H.). S.J.S. is an American Cancer Society Professor, and D.J.G.R. was the recipient of a NATO/SERC postdoctoral fellowship.

TABLE 1 Kinetic constants for the hydrolysis of the monomeric substrate *rac* 1,2-dihexanoyldithioleicthrin by wild-type and two mutant phospholipases A_2^*

Enzyme	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($s^{-1}M^{-1}$)
Wild-type	0.72	0.80	900
$\Delta 62-66$	0.88	0.50	1,760
$\Delta 62-66, W3F, L31W$	0.57	0.17	3,350

* For details see Volwerk *et al.*¹⁸. The $\Delta 62-66$ mutant has been described¹²; the $\Delta 62-66, W3F, L31W$ is the mutant used here.

Escherichia coli K12 strain PC2494 ($\Delta(laq-pro)$, supE, thi/F, traD36, proA⁺B⁺, lac Iq, LacZ Δ M15 (Phabagen Collection, Utrecht)) was used for plasmid construction and as a host for M13-derived vectors. Strain HB2154 (ara, $\Delta(lac-pro)$, thi/F⁺, pro⁺B⁺, lac Iq, LacZ Δ M25, mutL::Tn10)¹⁹ was used as a recipient strain in the mutagenesis experiments. As mutagenesis primers for site-directed mutagenesis we used: for W3F, 5'T.AAT.CAT.GCT.T*CG.AAA.CTG.A*A*A.TAA.TGC.CC3'; and for L31W, 5'C.TGA.TCC.ACC.C*C*A*.GCC.ACA.GTA.G3'. The sites of mutation in the sequence of the oligonucleotides, which were synthesized on a Bioscience 6800 DNA synthesizer, are denoted by an asterisk. Substitutions in the deletion mutant ($\Delta 62-66$, D59S, S60G, N67Y)¹² were introduced by the gapped duplex method, using amber selection²⁰. The resulting mutant pro-PLA₂ complementary DNA was sequenced by the dideoxy chain termination method²¹ after plaque-purifying. The sequenced gene was cloned into the expression vector, expressed in *E. coli* K12 strain MC4100, and the mutant phospholipase A_2 isolated and purified as described¹².

molecules per asymmetric unit), and its structure solved and refined at 2.4 Å resolution to an *R*-factor of 18.9% (for details see the legend to Fig. 2). The density for the inhibitor was clear and easy to interpret (Fig. 2). The inhibitor is bound to the calcium ion in the active site of the enzyme by both its phosphate group and the carbonyl oxygen of its amide bond (Fig. 3). These two ligands replace two water molecules that are the normal ligands of the calcium ion in the native structure⁷. The hydroxyl group of Tyr 69 makes a hydrogen bond with one of the oxygen atoms of the phosphate group. This agrees with site-directed

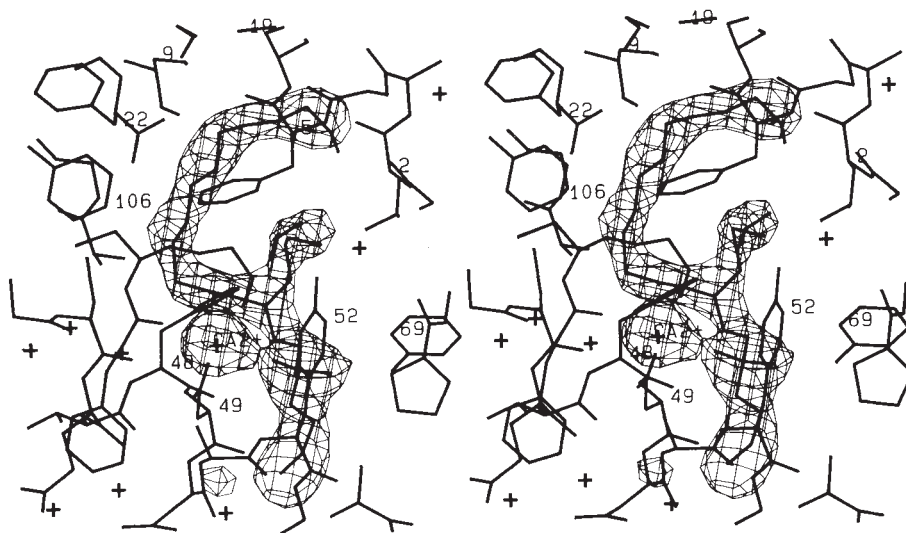
mutagenesis experiments, which indicated that Tyr 69 is important for the precise positioning of the phospholipid substrate in the enzyme's active site¹³. The acyl chain of the inhibitor makes extensive hydrophobic contacts with the disulphide bridge between residues 29 and 45, and with the side chains of residues Leu 2, Phe 5, Ile 9, Leu 19, Phe 22 and Tyr 52. The nitrogen of the amide bond forms a strong hydrogen bond with the N δ 1 atom of His 48. The short alkyl chain has no interactions closer than 4 Å with the protein. The glycol headgroup makes a hydrogen bond to the backbone carbonyl oxygen of Asp 49. All residues involved in binding are either identical or substituted with an equivalent functional group in the human cellular phospholipase A_2 (Fig. 3a).

The conformation of this inhibitor in the complex is similar to that of natural phospholipids as determined by X-ray crystallography and NMR in solution^{14,15}. As was also found in those studies, the 2-acyl chain has a kink of $\sim 90^\circ$ at the α -carbon atom of the fatty acid.

The active site residues of phospholipase A_2 (ref. 16) do not, within error, change in position on binding the inhibitor. Two water molecules that are near the N δ 1 atom of His 48 in the native structure however, have been replaced by the bound inhibitor. It has been proposed that one of these water molecules is activated by the His 48-Asp 99 couple and acts as the nucleophile attacking the carbonyl carbon atom of the scissile bond. The oxy-anion intermediate would then be stabilized by the calcium ion and the peptide NH group of Gly 30 (refs 16, 17). Indeed, in our structure the inhibitor's carbonyl oxygen atom is bound to the calcium ion and is at 2.7 Å from the NH group of Gly 30. But a strong hydrogen bond exists between the NH of the inhibitor's amide bond and the N δ 1 atom of His 48. In a true substrate-enzyme complex, the ester oxygen cannot form such a hydrogen bond with the unprotonated N δ 1 atom, which is the catalytically active species¹⁷. This extra hydrogen bond can contribute to the higher affinity of the enzyme for the inhibitor compared with that for the natural substrate.

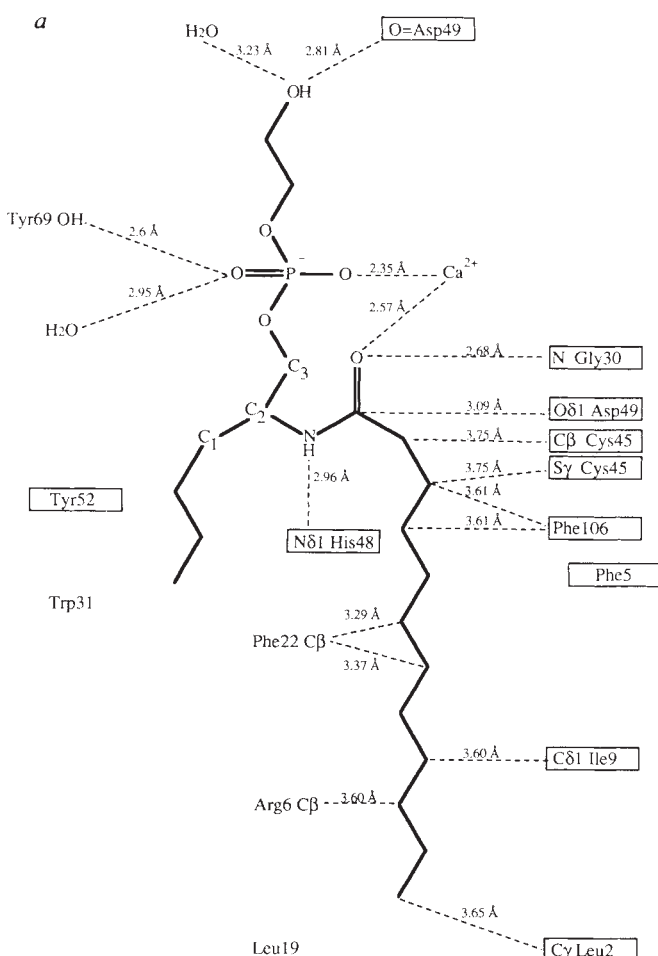
The absence of a water molecule in the active site close to the acyl-amido linkage calls into question the attacking nucleophile. The O δ 1 atom of Asp 49 is in the inhibitor-enzyme

FIG. 2 Stereoview of the inhibitor in ($mF_o - DF_c$), α_C (ref. 23) electron density. The calcium ion and the inhibitor molecule were excluded from the calculation of F_c and α_C . The mutant phospholipase A_2 complexed with the inhibitor was crystallized in a hanging drop set-up from 40% 2-methyl-2,4-pentanediol in 100 mM Tris(hydroxymethyl)aminomethane/HCl buffer, pH 7.9, with 5 mM $CaCl_2$ and 1 mg ml⁻¹ inhibitor. The space group was $P2_12_12_1$ with unit cell dimensions $a=52.57$ Å, $b=62.44$ Å and $c=85.58$ Å. There are two phospholipase A_2 molecules per asymmetric unit. X-ray diffraction data to 2.4 Å resolution were collected from one single crystal (0.7 mm $\times$ 0.3 mm $\times$ 0.15 mm) on a fast area-sensitive television detector system (Enraf-Nonius, Delft, The Netherlands). The data set was 89.1% complete to a resolution of 2.4 Å. The three-dimensional structure of the mutant phospholipase A_2 was solved by molecular replacement methods (rotation²⁴ and translation²⁵ functions) with the $\Delta 62-66$ mutant phospholipase A_2 dimer as the starting structure¹². The position and the orientation of both molecules were refined with the rigid body refinement option of the TNT-refinement program package²⁶. The same package was used for the subsequent crystallographic refinement. The structure of the inhibitor was built using the program BIOGRAF (Bioscience Corporation, Pasadena) and

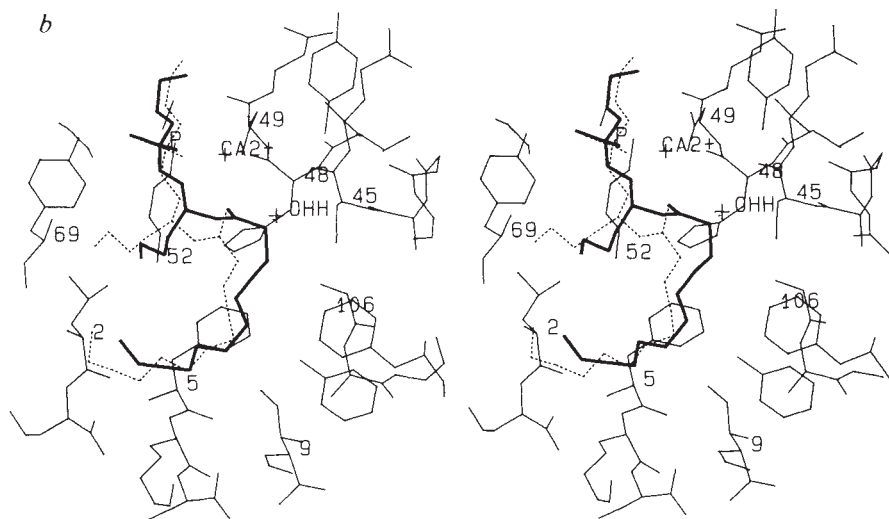


was fitted into density after the rigid body refinement. The quality of the model was analysed and improved at intervals during the refinement with the molecular graphics program FRODO (ref. 27) running on an Evans & Sutherland PS390 picture system. The final crystallographic *R*-factor is 0.189 for all data between 7 and 2.4 Å resolution, with tight restraints on the molecular geometry (for example, the r.m.s. deviation from ideal bond lengths is 0.012 Å).

FIG. 3 *a*, Schematic drawing of the interactions of the inhibitor with the enzyme. Residues within 5 Å distance from the inhibitor are drawn. Residues identical in the present mutant and cellular phospholipase A_2 are enclosed in boxes; the other residues are substituted as follows: Arg 6 $\rightarrow$ His, Leu 19 $\rightarrow$ Ala, Phe 22 $\rightarrow$ Tyr, Trp 31 $\rightarrow$ Val, Tyr 69 $\rightarrow$ Lys. *b*, Stereo view of the inhibitor (heavy lines) and a modelled substrate (dashed lines) in the active site. The substrate was obtained by changing the inhibitor's amide nitrogen atom into an ester oxygen atom and adjusting its conformation by hand to create space for the putative water molecule (indicated as OHH). The atomic positions of the water molecule and substrate were subjected to energy minimization while the protein environment was kept frozen. Model building and energy minimization were performed using the program BIOGRAF (Biodesign Corporation, Pasadena).



complex at 3.0 Å from the amide carbonyl carbon of the inhibitor, and we cannot completely exclude the possibility that this is the nucleophile. After nucleophilic attack an anhydride is formed as a covalently bound intermediate. It is unclear from this mechanism which group will donate a proton to the leaving group, because His 48 is unprotonated and no other side groups are available. Furthermore, in this mechanism there seems to be no role for the Asp 99–His 48 couple which is conserved in all phospholipases A_2 . Using model building, we tried replacing the amide bond of the inhibitor with an ester bond (see legend



to Fig. 3*b*) to see how a true substrate might bind in the active site, and whether enough space would become available for a water molecule. We found that a simple tilting of the ester group away from His 48 creates sufficient space for a water molecule (Fig. 3*b*); no perturbations of the enzyme were needed. The modelled substrate is liganded to the calcium ion both by its phosphate group and the carbonyl oxygen atom. The modelled water is hydrogen-bonded to the His 48 Nδ1 atom, and is at about 2.8 Å from the carbonyl carbon atom of the substrate. With natural substrates a water molecule could therefore act as the nucleophile in accordance with earlier proposals¹⁷.

The structure we describe of a complex of a lipolytic enzyme with a substrate analogue could be a useful starting point for the design of anti-inflammatory drugs, although the implications for the enzyme's mechanism are not yet fully understood.

Note added in proof: The coordinates for this phospholipase A_2 –inhibitor complex have been deposited with the protein databank, Chemistry Department, Brookhaven National Laboratory, Upton, New York 11973, USA (reference number 5P2P).

Received 2 May; accepted 22 August 1990.

- Hirata, F. & Axelrod, J. *Science* **209**, 1082–1090 (1980).
- Vadas, P. & Pruzanski, W. *Lab. Invest.* **4**, 391–404 (1986).
- Flower, R. *Tr. pharmacol. Sci.* **2**, 186–188 (1981).
- Seilhamer, J. J. *et al. J. biol. Chem.* **264**, 5335–5338 (1989).
- Kramer, R. M. *et al. J. biol. Chem.* **264**, 5768–5775 (1989).
- Mizushima, H. *et al. J. Biochem.* **105**, 520–525 (1989).
- Dijkstra, B. W., Kalk, K. H., Hol, W. G. J. & Drenth, J. *J. molec. Biol.* **147**, 93–123 (1981).
- Dijkstra, B. W., Renetseder, R., Kalk, K. H., Hol, W. G. J. & Drenth, J. *J. molec. Biol.* **168**, 163–179 (1983).
- Brunie, S., Bolin, J., Gewirth, P. & Sigler, P. B. *J. biol. Chem.* **260**, 9742–9747 (1985).
- Renetseder, R., Brunie, S., Dijkstra, B. W., Drenth, J. & Sigler, P. B. *J. biol. Chem.* **260**, 11627–11634 (1985).
- de Haas, G. H., Dijkman, R., van Oort, M. G. & Verger, R. *Biochim. biophys. Acta* **1043**, 75–82 (1990).
- Kuipers, O. P. *et al. Science* **244**, 82–85 (1989).
- Kuipers, O. P., Dijkman, R., Pais, C. E. G. M., Verheij, H. M. & de Haas, G. H. *Protein Engng* **2**, 467–471 (1989).
- Hitchcock, P. B., Mason, R., Thomas, K. M. & Shipley, G. G. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3036–3040 (1974).
- Seelig, J. & Browning, J. L. *FEBS Lett.* **92**, 41–44 (1978).
- Dijkstra, B. W., Kalk, K. H. & Drenth, J. *Nature* **289**, 604–606 (1981).
- Verheij, H. M. *et al. Biochemistry* **19**, 743–750 (1980).
- Volwerk, J. J., Dedieu, A. G. R., Verheij, H. M., Dijkman, R. & de Haas, G. H. *Recl. Trav. Chim. Pays-Bas* **98**, 214–220 (1979).
- Carter, P., Bedouelle, H. & Winter, G. *Nucleic Acids Res.* **13**, 4431–4443 (1985).
- Kramer, W. *et al. Nucleic Acids Res.* **12**, 9441–9456 (1984).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Dijkman, R., Dekker, N. & de Haas, G. H. *Biochim. biophys. Acta* **1043**, 67–74 (1990).
- Read, R. *Acta crystallogr. Sect. A* **42**, 140–149 (1986).
- Crowther, R. A. *In The Molecular Replacement Method* (ed. Rossman, M. G.) 173–178 (Gordon & Breach, New York, 1972).
- Crowther, R. A. & Blow, D. M. *Acta crystallogr.* **23**, 544–548 (1967).
- Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. *Acta crystallogr. Sect. A* **43**, 489–501 (1987).
- Jones, T. A. *Meth. Enzym.* **115**, 157–171 (1985).

Mandelate racemase and muconate lactonizing enzyme are mechanistically distinct and structurally homologous

David J. Neidhart^{*†}, George L. Kenyon[‡], John A. Gerlt[§] & Gregory A. Petsko^{*†}

^{*} Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[†] Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143, USA

[‡] Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742, USA

[§] Present addresses: Protein Crystallography Laboratory, D-47E, Abbott Laboratories, Abbott Park, Illinois 60064, USA (D.J.N.); Rosenstiel Center, Brandeis University, Waltham, Massachusetts 02254, USA (G.A.P.)

MANDELATE racemase (MR) and muconate lactonizing enzyme (MLE) catalyse separate and mechanistically distinct reactions necessary for the catabolism of aromatic acids by *Pseudomonas putida*^{1–3}. The X-ray crystal structure of MR, solved at 2.5 Å resolution, reveals that the secondary, tertiary and quaternary structures of MR and MLE⁴ are remarkably similar; also, MR and MLE are about 26% identical in primary structure⁵. However, MR has no detectable MLE activity and vice versa. Thus, MR and MLE constitute the first example of enzymes that catalyse different reactions, as opposed to mechanistically identical reactions on different substrates, yet possess sufficient structural and sequence identity that they are likely to have evolved from a

common ancestor. The discovery that MR and MLE catalyse different reactions but share a common structural framework has broad implications for the natural evolution of enzymes and metabolic pathways, as well as for the rational modification of enzyme activities.

We were fortunate to discover a crystal form of MR that contained only a single polypeptide chain in the crystallographic asymmetric unit as this greatly expedited solution of the structure by multiple isomorphous replacement (Fig. 1). The crystal structure reveals that MR is tightly packed as an octamer of identical subunits possessing 422 (D_4) point symmetry. This subunit arrangement has previously been observed in crystal structures of only three proteins: methemerythrin⁶, glycolate oxidase⁷ and muconate lactonizing enzyme⁴. Presumably, the fact that MR and MLE share a common quaternary structure reflects the striking similarity of tertiary structure between the subunits of the two enzymes. Both enzymes possess a parallel α/β barrel domain similar to that first described in triosephosphate isomerase⁸ and subsequently in 16 other enzymes⁹, and their polypeptide chain folds are virtually identical (Fig. 2).

A detailed analysis of the structural homology between MR and MLE has been performed using the method of Rossmann and Argos¹⁰. This analysis attempts to identify the best set of 'structurally equivalent' residues between two protein structures on the basis of the relative positions of $C\alpha$ atoms and their order in the sequence; residues are accepted as equivalent only if their positions are equivalent within certain statistical criteria. Of the 358 amino-acid residues in a single subunit of MR, 214 (60%) were found to be structurally equivalent to residues in MLE, with a root-mean-square distance of 1.3 Å between equivalent $C\alpha$ atoms within these sets (Fig. 2). For comparison, 60% of the residues in different haem-binding proteins are structurally equivalent, and in pyridine nucleotide dehydrogenases, 64% of the residues are equivalent¹⁰.

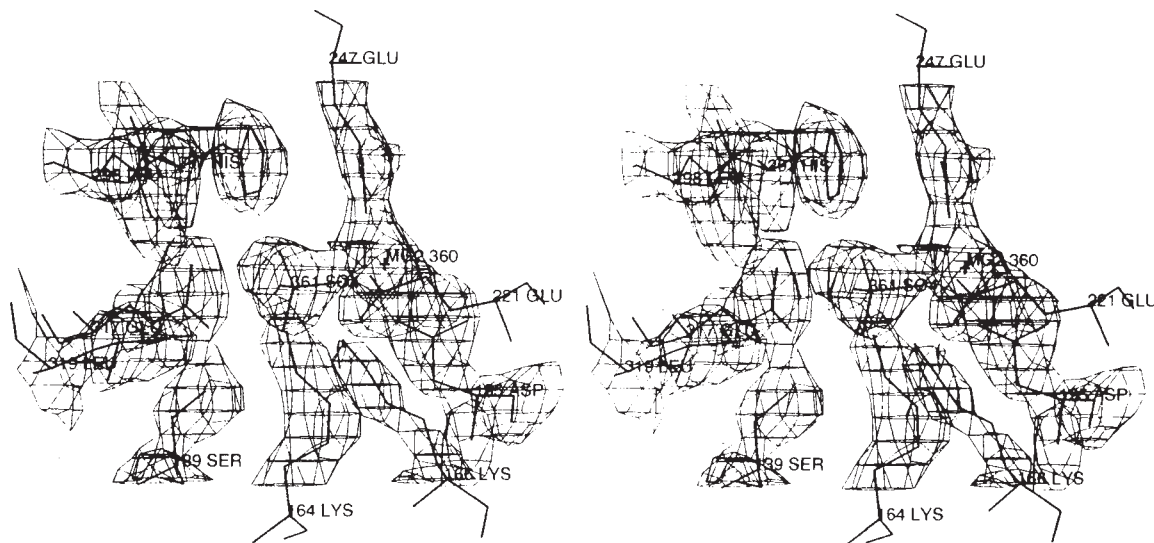


FIG. 1 Electron density and atomic model of mandelate racemase in the active site region, displayed in stereo.

METHODS. Mandelate racemase was overproduced and purified as previously described¹⁷. We have previously reported two crystal forms of mandelate racemase grown from polyethylene glycol¹⁸; however, the crystal structure was solved using a new crystal form grown from ammonium sulphate. This form (form III) contains only one subunit of mandelate racemase in the crystallographic asymmetric unit whereas previous forms contained two identical subunits in the asymmetric unit. Form III crystals were grown from 20–25% ammonium sulphate by vapour diffusion, and exhibit the symmetry of the tetragonal space group $I422$, with unit cell dimensions $a = b = 124.5$ Å, $c = 106.2$ Å. Diffraction data were collected by conventional diffractometry and the structure was solved by multiple isomorphous replacement using four heavy-atom derivatives (Table 1). A unique feature of the structure determination was the use of a diplatinum heavy

atom reagent, PIP (di- μ -iodo-bis(ethylenediamine)diplatinum II nitrate) previously reported in reference to the low-resolution crystal structure of the nucleosome core particle¹⁹. Application of phase improvement algorithms²⁰ to the initial multiple isomorphous replacement phases improved the overall figure of merit from 0.64 to 0.85 at 3.0 Å resolution and permitted an unambiguous tracing of the polypeptide chain. The crystal structure has subsequently been refined to an R -factor of 18.3% for native data between 5.0 and 2.5 Å resolution using molecular dynamics refinement (program XPLOR; ref. 21) and restrained least squares (program PROLSQ; ref. 22). The refined model is displayed in stereo, along with a grid representation of an electron density map calculated using coefficients $2|F_{\text{observed}}| - |F_{\text{calculated}}|$ and calculated phases. A complete description of the crystallization, structure solution and refined atomic model of MR will be published separately.

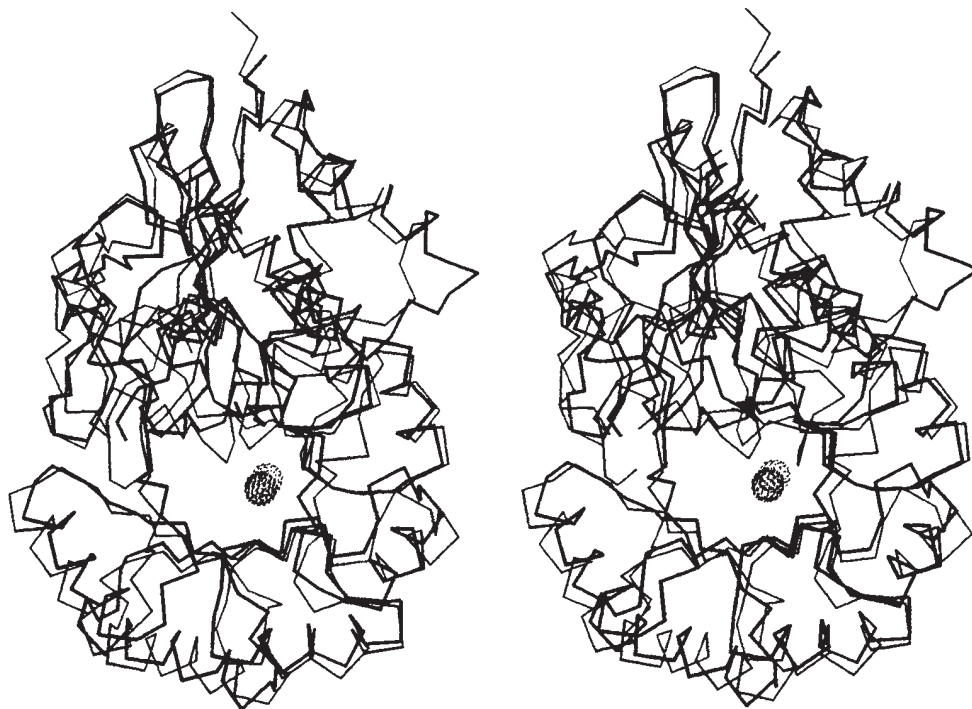
TABLE 1 Heavy atom parameters after difference Patterson refinement²³

Derivative	Resolution	Site	x	y	z	Occ.*	B (Å ²)
Europium chloride (1 mM, 24 h)	3 Å	1	0.285	0.467	0.163	0.145	28
		2	0.373	0.455	0.097	0.067	80
Pt(NH ₃) ₂ Cl ₂ (1 mM, 24 h)	4 Å	1	0.384	0.195	0.124	0.155	85
PIP (1 mM, 24 h)	3.5 Å	1	0.375	0.195	0.122	0.157	20
		2	0.435	0.174	0.113	0.131	20
Mercuric chloride (0.5 mM, 30 min)	3 Å	1	0.461	0.284	0.102	0.102	14
		2	0.373	0.447	0.108	0.133	26

Derivatives were prepared by transferring crystals to a buffered (pH 6.0) synthetic mother liquor containing the heavy atom reagent. PIP is a diplatinum heavy atom reagent (di- μ -iodo-bis(ethylene diamine) diplatinum II nitrate) (see Fig. 1 legend).

* Occupancy in arbitrary units.

FIG. 2 Stereo diagram of C α backbone tracing of mandelate racemase with comparison to muconate lactonizing enzyme. The backbone of MR is displayed in bold lines, with the backbone of MLE superposed in thinner lines. Note that all regular elements of secondary structure are conserved between the two enzymes.



Among the 214 structurally equivalent residues in MR and MLE, 64 pairs (30%) are chemically identical and a complete alignment of the sequences of the two enzymes based on this 'core' alignment reveals about 30% overall sequence identity (data not shown). It is believed that a structurally based sequence alignment is more meaningful for distantly related proteins than an alignment based solely on sequence, because the three-dimensional structure of related proteins is more rigorously preserved than either DNA or amino-acid sequence¹¹⁻¹³. This

level of sequence identity is very unlikely to occur by chance¹⁴ and, in combination with the observed structural homology, suggests that MR and MLE are derived from a common precursor.

Given this evidence for divergent evolution, it is not surprising that the active sites of MR and MLE share important structural features. For instance, both enzymes require a divalent metal ion for activity and the crystallographically determined positions of these ions superpose to within 1 Å (Fig. 2).

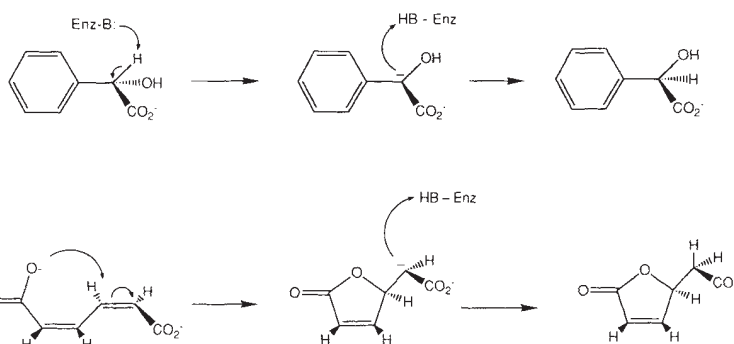


FIG. 3 Reactions catalysed by mandelate racemase (top) and muconate lactonizing enzyme (bottom).

Numerous cases are known where enzymes with different substrate specificities have arisen by divergent evolution from an ancestral enzyme, for instance malate and lactate dehydrogenases, or chymotrypsin and trypsin. To our knowledge, however, MR and MLE constitute the first case where different chemical reactions (Fig. 3) are catalysed by two enzymes that can be demonstrated to have evolved from a common ancestor. MR catalyses the interconversion of two enantiomers, whereas MLE catalyses the cycloisomerization of *cis,cis*-muconic acid to muconolactone by *syn* addition to a double bond. Thus, the process of divergent evolution can rework the catalytic apparatus of an enzyme without resorting to large-scale rearrangements of the protein framework and with significant conservation of sequence. This result has obvious implications with respect to rational redesign of existing enzymes.

A thorough explanation of how such an alteration of enzyme reaction mechanism occurred must await complete descriptions of the catalytic mechanisms of MR and MLE. However, two similarities in the reactions catalysed by MR and MLE may be noted: (1) both may proceed through formation of an intermediate carbanion located alpha to a carboxyl group^{15,16}, and (2) both involve subsequent protonation of this putative carbanion (Fig. 3). In MR the carbanion is generated by deprotonation of the α carbon, whereas in MLE the carbanion is generated by attack of the remote carboxylate group on a double bond. Both enzymes require a divalent metal ion for activity but use no other cofactor; MR is most active with Mg^{2+} bound as cofactor, whereas Mn^{2+} is preferred for MLE, but either enzyme can use either ion. Given the nature of the carbanionic intermediates, it is likely that a principal role of the metal ion in each enzyme is to polarize the carboxyl group of substrate.

As MR is found only in a subset of those strains of *Pseudomonas* that can metabolize catechol, MR might have arisen as the result of duplication of the MLE structural gene followed by random differentiation. Once the new gene product acquired a useful and unique catalytic activity, phenotypic selection eventually produced an efficient and selective enzyme. In the case of MR and MLE, this final selection process has apparently 'covered the tracks' of the differentiation process; MR has no detectable MLE activity and vice versa (W. Pieken, K. Taylor and J. Kozarich, personal communication). It may, however, be possible to retrace the steps of the evolutionary process under controlled conditions. Efforts are underway to generate MR activity from MLE by site-directed and random mutagenesis of the MLE gene. □

Received 25 June; accepted 24 August 1990.

- Halpin, R. A., Hegeman, G. D. & Kenyon, G. L. *Biochemistry* **20**, 1525-1533 (1981).
- Hegeman, G. D. *J. Bact.* **91**, 1140-1154 (1966).
- Ornston, L. N. *J. biol. Chem.* **241**, 3800-3810 (1966).
- Goldman, A., Ollis, D. L. & Steitz, T. A. *J. molec. Biol.* **194**, 143-153 (1987).
- Tsou, A. Y. *et al. Biochemistry* (in the press).
- Stenkamp, R. E. & Jensen, L. H. In *Structural Aspects of Recognition and Assembly in Biological Molecules* (eds Balaban, M., Sussman, J. L., Traub, W. & Yonath, A.) (Balaban ISS, Philadelphia, 1981).
- Lindqvist, Y. & Brändén, C.-L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6855-6859 (1985).
- Banner, D. W. *et al. Nature* **255**, 609-614 (1975).
- Farber, G. K. & Petsko, G. A. *Trends biochem. Sci.* **15**, 228-234 (1990).
- Rossmann, M. J. & Argos, P. *J. biol. Chem.* **250**, 7525-7532 (1975).
- Greer, J. *J. molec. Biol.* **153**, 1027-1042 (1981).
- Schultz, G. E. & Schirmer, R. H. *Principles of Protein Structure* (Springer, New York, 1979).
- Chothia, C. & Lesk, A. M. *EMBO J.* **5**, 823-826 (1986).
- Schultz, G. E. *Angew. Chem. Int. Ed. Engl.* **20**, 143-151 (1981).
- Lin, D. T. *et al. J. Am. chem. Soc.* **110**, 323-324 (1988).
- Ngai, K.-L. & Kallen, R. G. *Biochemistry* **22**, 5231-5236 (1983).
- Tsou, A. Y., Ransom, S. C., Gerlt, J. A., Powers, V. M. & Kenyon, G. L. *Biochemistry* **28**, 969-975 (1989).
- Neidhart, D. J. *et al. J. biol. Chem.* **263**, 9268-9270 (1988).
- O'Halloran, T. V., Lippard, S. J., Richmond, T. J. & Klug, A. *J. molec. Biol.* **194**, 705-712 (1987).
- Wang, B. C. *Meth. Enzym.* **115**, 90-111 (1985).
- Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458-460 (1987).
- Hendrickson, W. A. & Konnert, J. H. in *Biomolecular Structure, Function, Conformation and Evolution* (ed. Srinivasan, R.) Vol. 1 43-57 (Pergamon, Oxford, 1980).
- Terwilliger, T. C. & Eisenberg, D. *Acta. crystallogr.* **A39**, 813-817 (1983).

ACKNOWLEDGEMENTS. We thank P. L. Howell, S. Ransom, A. Tsou, V. Powers, W. Pieken, K. Taylor, and J. Landro for assistance. We also thank A. Goldman for providing atomic coordinates of MLE in advance of publication, and J. Kozarich for his continuing interest and assistance in these studies.

nature is available in microform.



University Microfilms

International reproduces this publication in microform: microfiche and 16mm or 35mm film. For information about this publication or any of the more than 13,000 titles we offer, complete and mail the coupon to: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106. Call us toll-free for an immediate response: 800-521-3044. Or call collect in Michigan, Alaska and Hawaii: 313-761-4700.

☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

City _____

State _____ Zip _____

Phone () _____

University
Microfilms
International